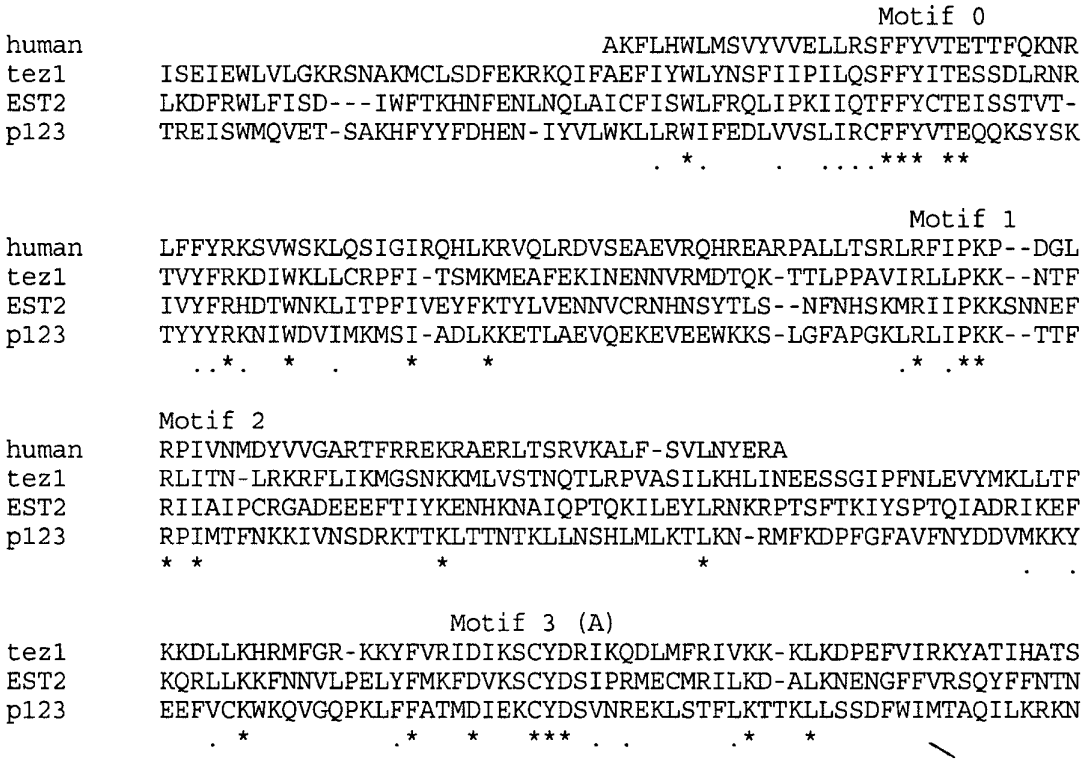

FIG. 1

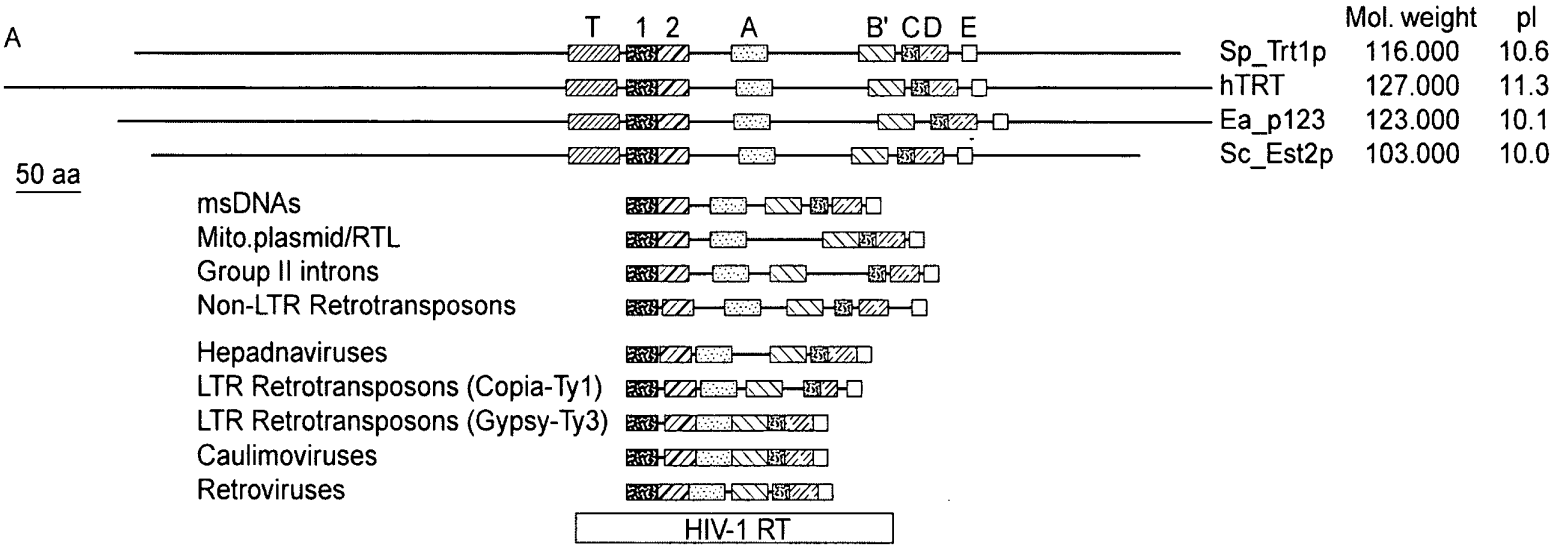
A
T 1 2 A B' CD E
Mol. weight pI
Sp_Trt1p 116.000 10.6
hTRT 127.000 11.3
Ea_p123 123.000 10.1
Sc_Est2p 103.000 10.0
50 aa
msDNAs
Mito.plasmid/RTL
Group II introns
Non-LTR Retrotransposons
Hepadnaviruses
LTR Retrotransposons (Copia-Ty1)
LTR Retrotransposons (Gypsy-Ty3)
Caulimoviruses
Retroviruses
HIV-1 RT

FIG. 2

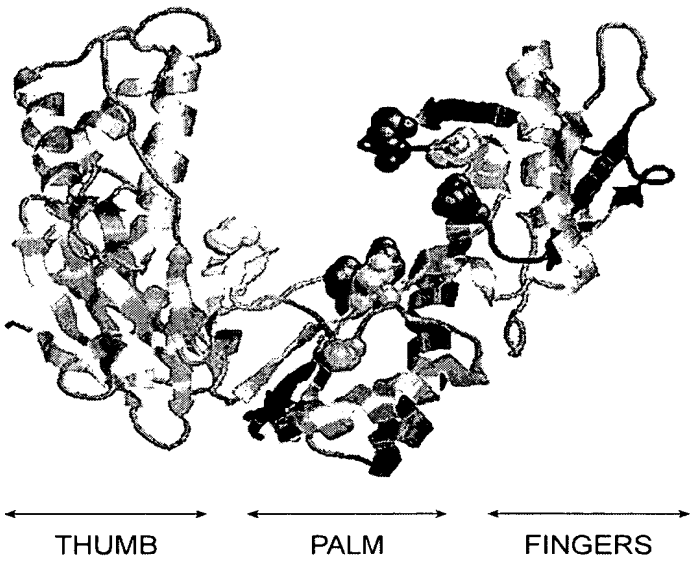
THUMB
PALM
FINGERS

FIG. 3

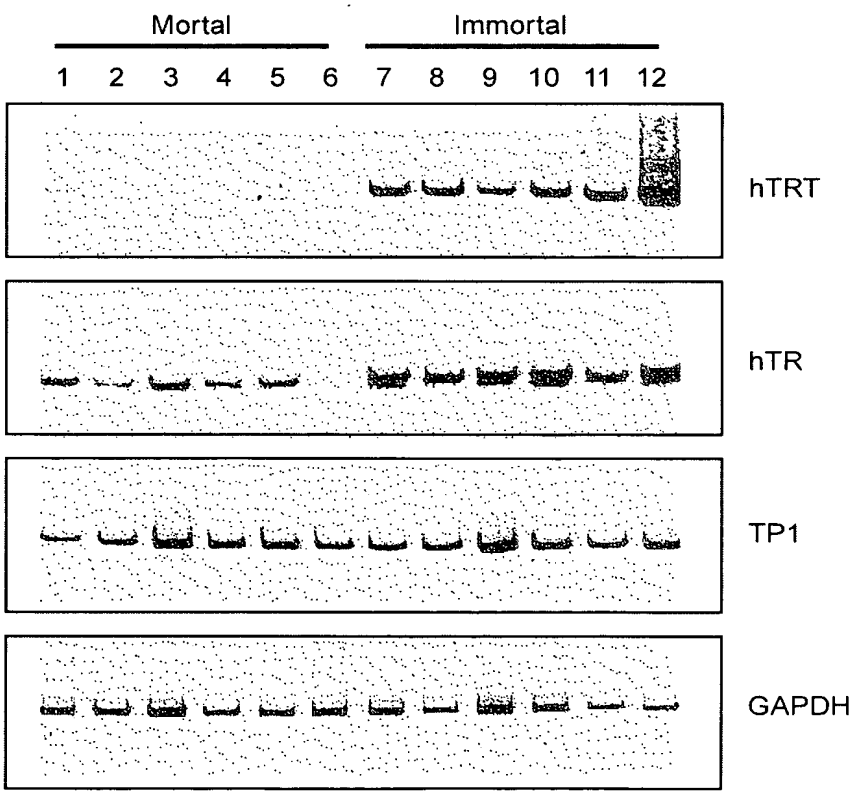
Mortal
Immortal
1 2 3 4 5 6 7 8 9 10 11 12
hTRT
hTR
TP1
GAPDH

FIG. 5

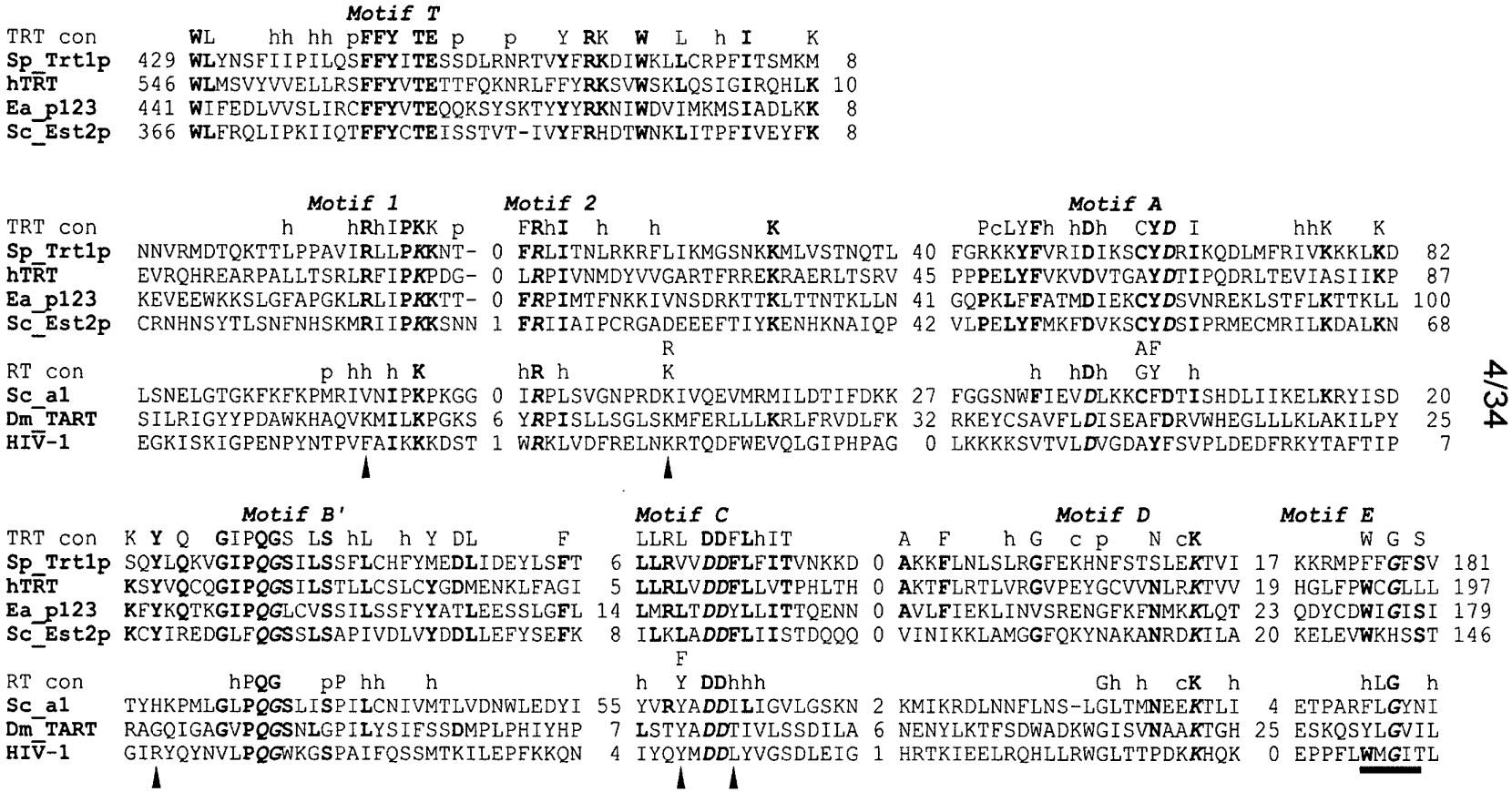
Motif T
TRT con WL hh hh pFFY TE p p Y RK W L h I K
Sp_Trt1p 429 WLYNSFIIPILQSFFYITESSDLRNRTVYFRKDIWKLLCRPFITSMKM 8
hTRT 546 WLMSVYVVELLRSFFYVTETTFQKNRLFFYRKSVWSKLQSIGIRQHLK 10
Ea_p123 441 WIFEDLVVSLIRCFFYVTEQQKSYSKTYYYRKNIWDVIMKMSIADLKK 8
Sc_Est2p 366 WLFRQLIPKIIQTFFYCTEISSTVT-IVYFRHDTWNKLITPFIVEYFK 8
Motif 1
Motif 2
Motif A
TRT con h hRhIPKK p FRhI h h K PcLYFh hDh CYD I hhK K
Sp_Trt1p NNVRMDTQKTTLPPAVIRLLPKKNT- 0 FRLITNLRKRFLIKMGSNKKMLVSTNQTL 40 FGRKKYFVRIDIKSCYDRIKQDLMFRIVKKKLKD 82
hTRT EVRQHREARPALLTSRLRFIPKPDG- 0 LRPIVNMDYVVGARTFRREKRAERLTSRV 45 PPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKP 87
Ea_p123 KEVEEWKKSLGFAPGKLRLIPKKTT- 0 FRPIMTFNKKIVNSDRKTTKLTTNTKLLN 41 GQPKLFFATMDIEKCYDSVNREKLSTFLKTTKLL 100
Sc_Est2p CRNHNSYTLSNFNHSKMRIIPKKSNN 1 FRIIAIPCRGADEEEFTIYKENHKNAIQP 42 VLPELYFMKFDVKSCYDSIPRMECMRILKDALKN 68
R
AF
RT con p hh h K hR h K h hDh GY h
Sc_a1 LSNELGTGKFKFKPMRIVNIPKPKGG 0 IRPLSVGNPRDKIVQEVMRMILDTIFDKK 27 FGGSNWFIEVDLKKCFDTISHDLIIKELKRYISD 20
Dm_TART SILRIGYYPDAWKHAQVKMILKPGKS 6 YRPISLLSGLSKMFERLLLKRLFRVDLFK 32 RKEYCSAVFLDISEAFDRVWHEGLLLKLAKILPY 25
HIV-1 EGKISKIGPENPYNTPVFAIKKKDST 1 WRKLVDFRELNKRTQDFWEVQLGIPHPAG 0 LKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIP 7
Motif B'
Motif C
Motif D
Motif E
TRT con K Y Q GIPQGS LS hL h Y DL F LLRL DDFLhIT A F h G c p N cK W G S
Sp_Trt1p SQYLQKVGIPQGSILSSFLCHFYMEDLIDEYLSFT 6 LLRVVDDFLFITVNKKD 0 AKKFLNLSLRGFEKHNFSTSLEKTVI 17 KKRMPFFGFSV 181
hTRT KSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAGI 5 LLRLVDDFLLVTPHLTH 0 AKTFLRTLVRGVPEYGCVVNLRKTVV 19 HGLFPWCGLLL 197
Ea_p123 KFYKQTKGIPQGLCVSSILSSFYYATLEESSLGFL 14 LMRLTDDYLLITTQENN 0 AVLFIEKLINVSRENGFKFNMKKLQT 23 QDYCDWIGISI 179
Sc_Est2p KCYIREDGLFQGSSLSAPIVDLVYDDLLEFYSEFK 8 ILKLADDFLIISTDQQQ 0 VINIKKLAMGGFQKYNAKANRDKILA 20 KELEVWKHSST 146
F
RT con hPQG pP hh h h Y DDhhh Gh h cK h hLG h
Sc_a1 TYHKPMLGLPQGSLISPILCNIVMTLVDNWLEDYI 55 YVRYADDILIGVLGSKN 2 KMIKRDLNNFLNS-LGLTMNEEKTLI 4 ETPARFLGYNI
Dm_TART RAGQIGAGVPQGSNLGPILYSIFSSDMPLPHIYHP 7 LSTYADDTIVLSSDILA 6 NENYLKTFSDWADKWGISVNAAKTGH 25 ESKQSYLGVIL
HIV-1 GIRYQYNVLPQGWKGSPAIFQSSMTKILEPFKKQN 4 IYQYMDDLYVGSDLEIG 1 HRTKIEELRQHLLRWGLTTPDKKHQK 0 EPPFLWMGITL

FIG. 4

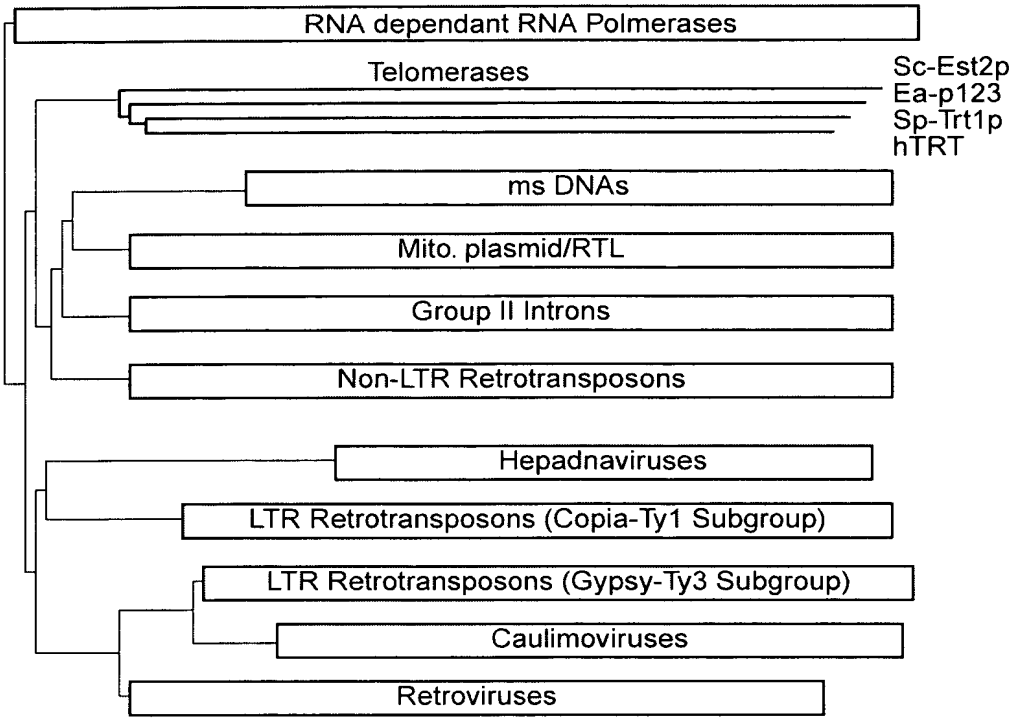
RNA dependant RNA Polmerases
Telomerases
Sc-Est2p
Ea-p123
Sp-Trt1p
hTRT
ms DNAs
Mito. plasmid/RTL
Group II Introns
Non-LTR Retrotransposons
Hepadnaviruses
LTR Retrotransposons (Copia-Ty1 Subgroup)
LTR Retrotransposons (Gypsy-Ty3 Subgroup)
Caulimoviruses
Retroviruses

FIG. 6

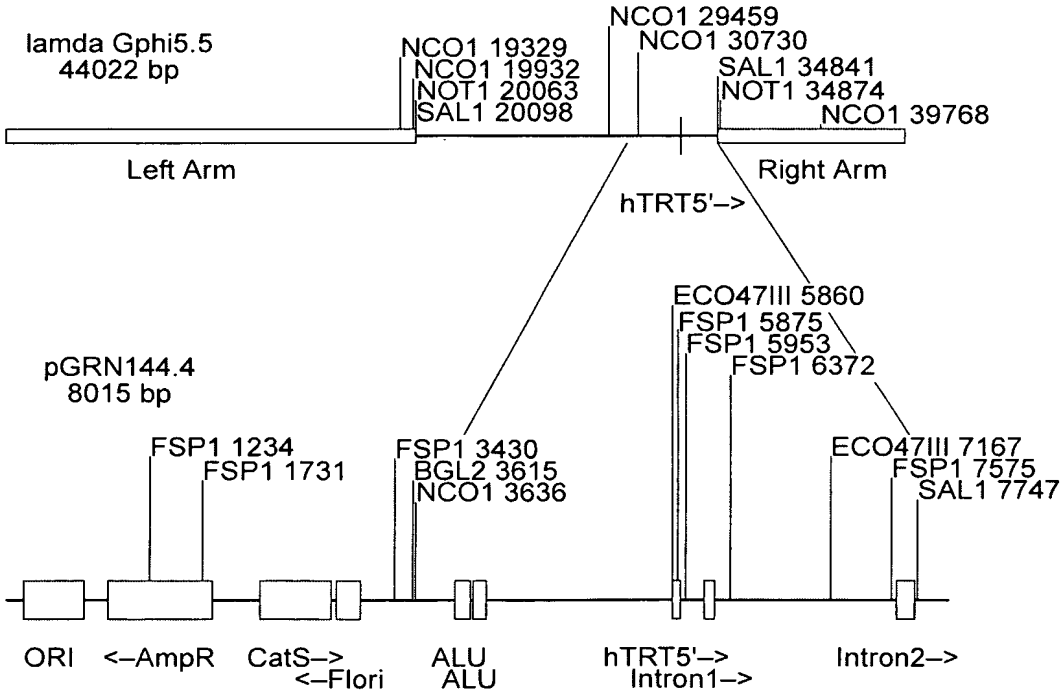
lamda Gphi5.5
44022 bp
NCO1 19329
NCO1 19932
NOT1 20063
SAL1 20098
NCO1 29459
NCO1 30730
SAL1 34841
NOT1 34874
NCO1 39768
Left Arm
Right Arm
hTRT5'->
ECO47III 5860
FSP1 5875
FSP1 5953
FSP1 6372
pGRN144.4
8015 bp
FSP1 1234
FSP1 1731
FSP1 3430
BGL2 3615
NCO1 3636
ECO47III 7167
FSP1 7575
SAL1 7747
ORI
<–AmpR
CatS–>
<–Flori
ALU
ALU
hTRT5'–>
Intron1–>
Intron2–>

FIG. 7

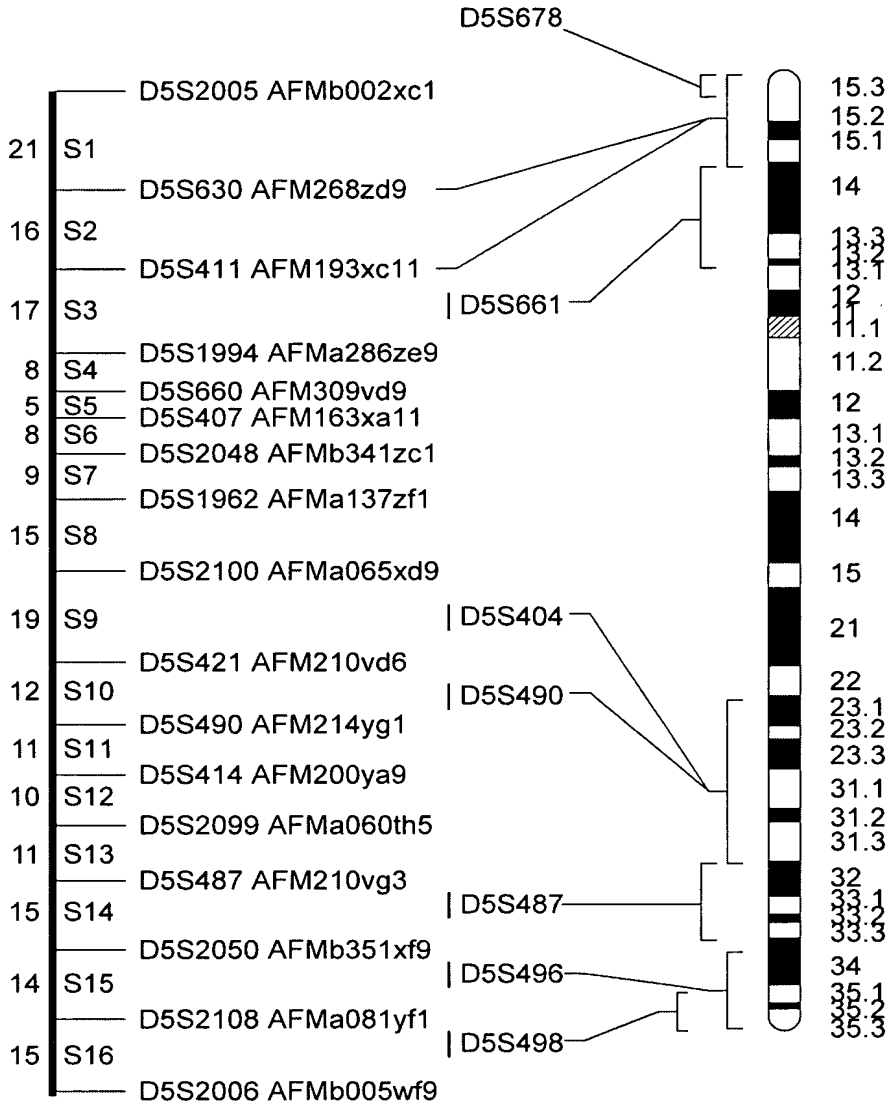
D5S678
D5S2005 AFMb002xc1
21 S1
D5S630 AFM268zd9
16 S2
D5S411 AFM193xc11
17 S3
D5S661
D5S1994 AFMa286ze9
8 S4
D5S660 AFM309vd9
5 S5
D5S407 AFM163xa11
8 S6
D5S2048 AFMb341zc1
9 S7
D5S1962 AFMa137zf1
15 S8
D5S2100 AFMa065xd9
19 S9
D5S404
D5S421 AFM210vd6
12 S10
D5S490
D5S490 AFM214yg1
11 S11
D5S414 AFM200ya9
10 S12
D5S2099 AFMa060th5
11 S13
D5S487 AFM210vg3
15 S14
D5S487
D5S2050 AFMb351xf9
14 S15
D5S496
D5S2108 AFMa081yf1
15 S16
D5S498
D5S2006 AFMb005wf9
15.3
15.2
15.1
14
13.3
13.2
13.1
12
11
11.1
11.2
12
13.1
13.2
13.3
14
15
21
22
23.1
23.2
23.3
31.1
31.2
31.3
32
33.1
33.2
33.3
34
35.1
35.2
35.3

FIG. 8

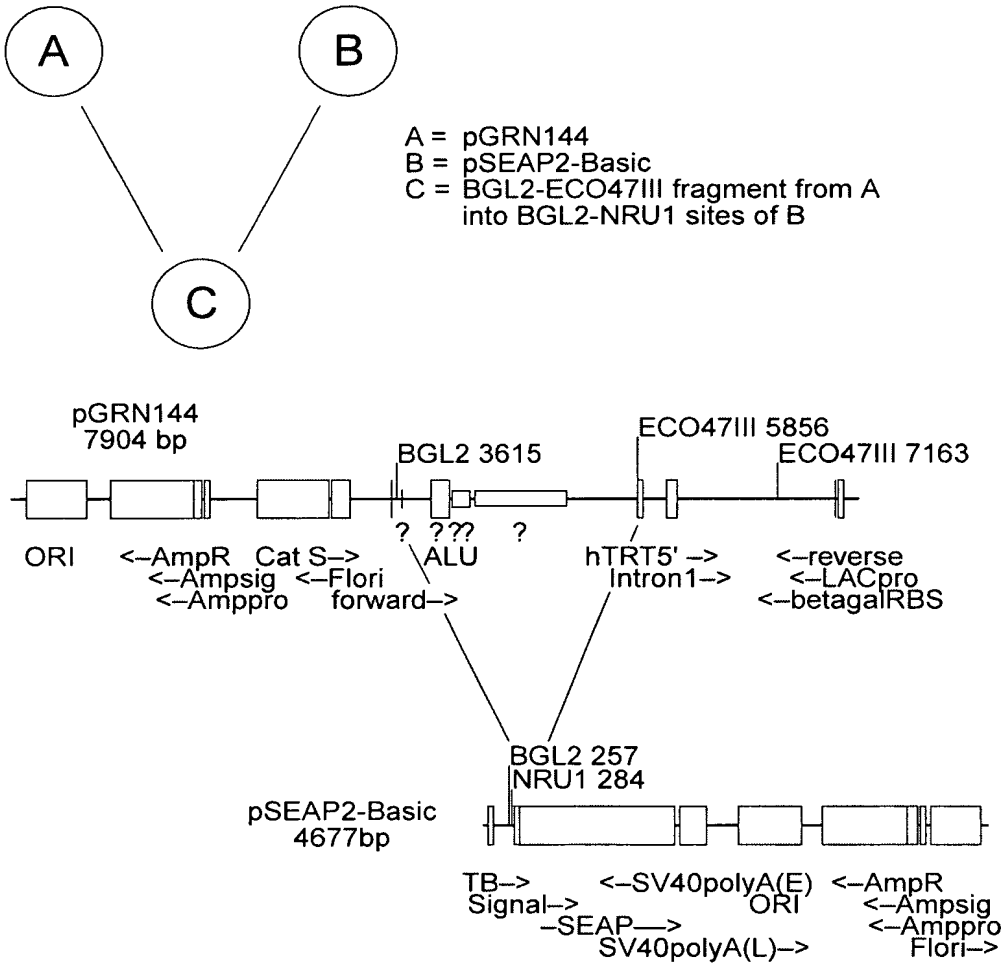
A
B
C
A = pGRN144
B = pSEAP2-Basic
C = BGL2-ECO47III fragment from A
into BGL2-NRU1 sites of B
pGRN144
7904 bp
BGL2 3615
ECO47III 5856
ECO47III 7163
? ??? ?
ORI
<–AmpR
<–Ampsig
<–Amppro
Cat S–>
<–Flori
forward–>
ALU
hTRT5' –>
Intron1–>
<–reverse
<–LACpro
<–betagalRBS
BGL2 257
NRU1 284
pSEAP2-Basic
4677bp
TB–>
Signal–>
–SEAP—>
<–SV40polyA(E)
ORI
SV40polyA(L)–>
<–AmpR
<–Ampsig
<–Amppro
Flori–>

FIG. 9

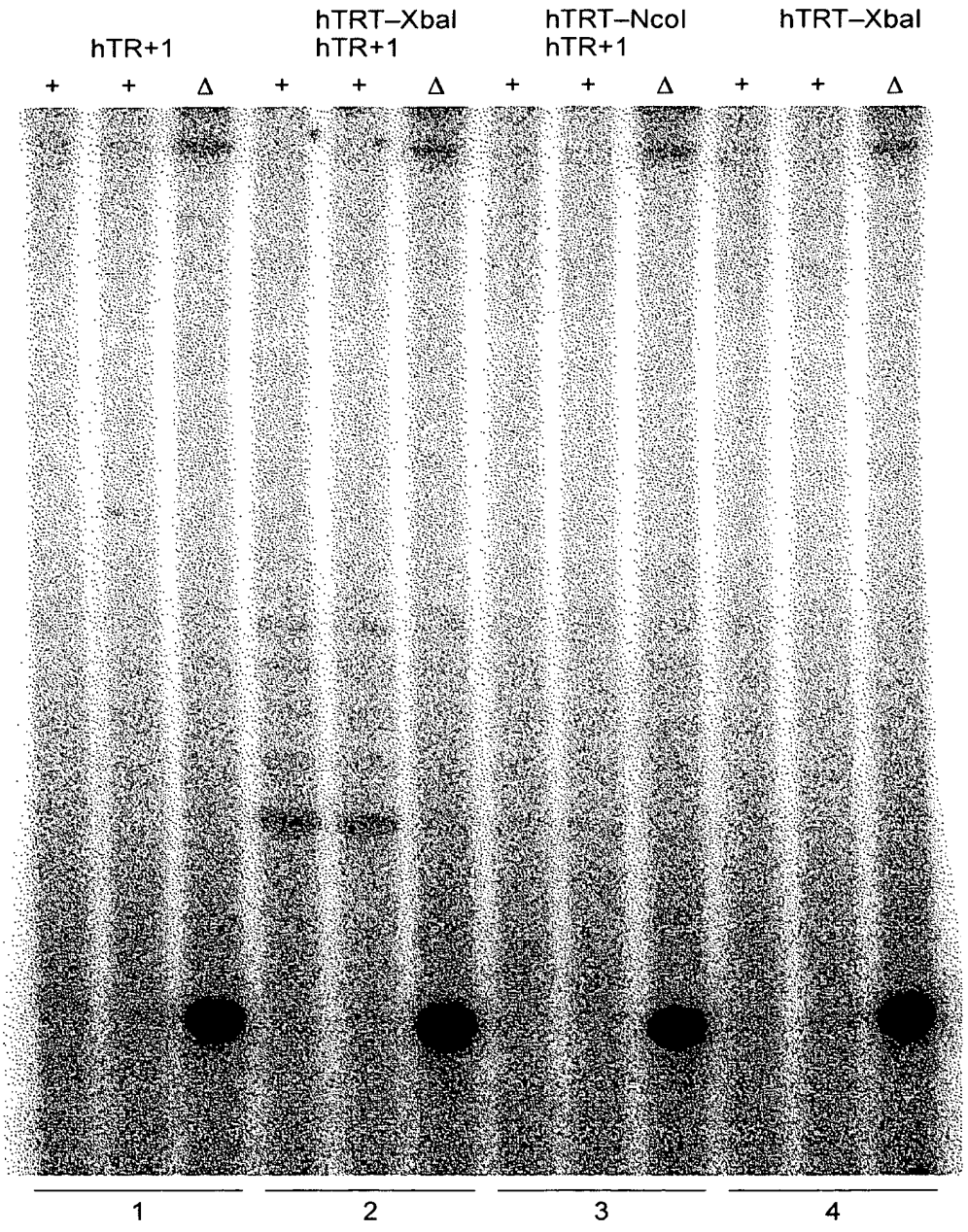

FIG. 10A

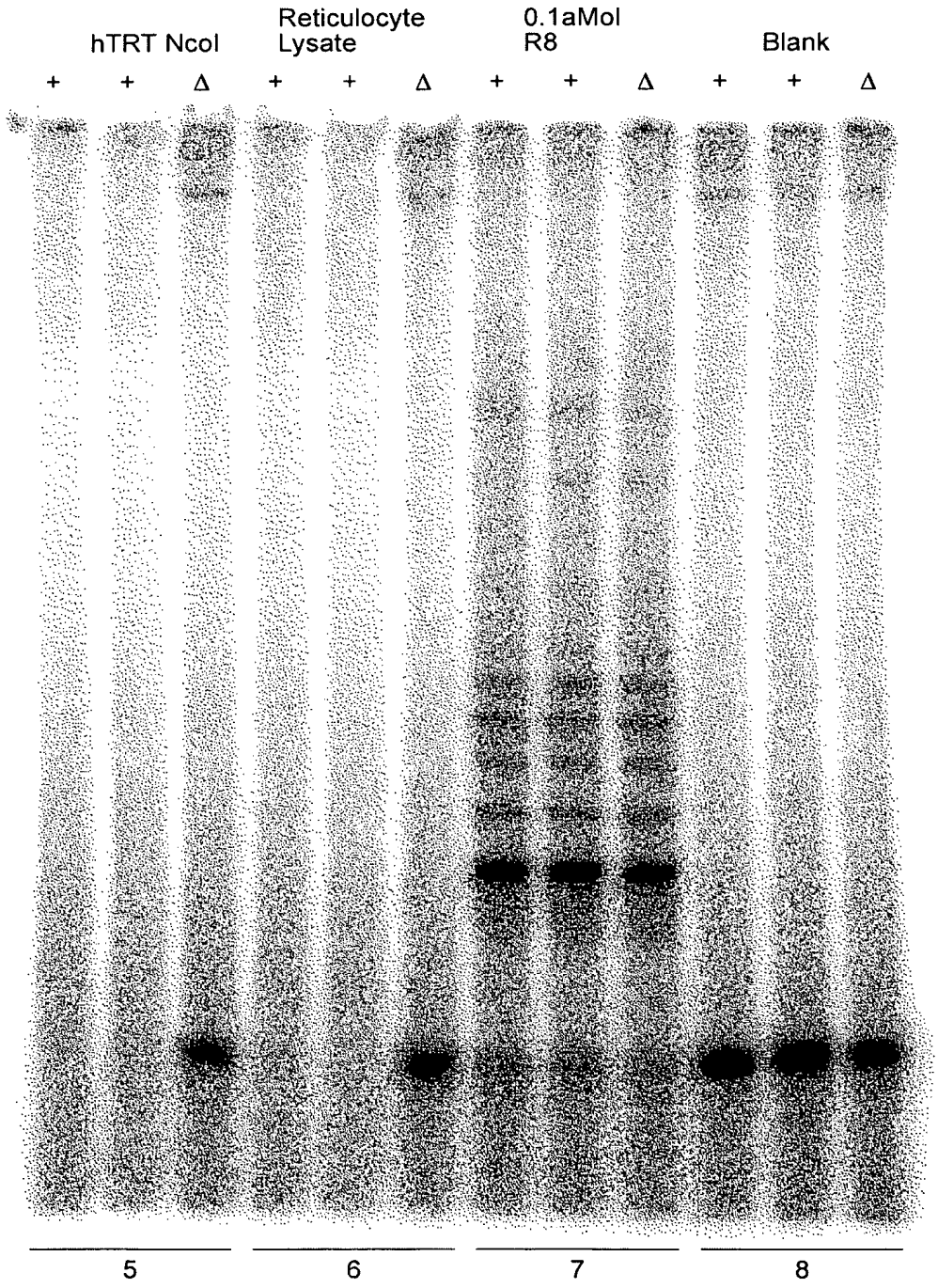

FIG. 10B

Telomerase Specific Motifs

		MOTIF T		MOTIF T'
TRT con		Wl FFY TE y Rk W l I		E V
hTRT	546	WLMSVYVVELLRSFFYVTETTFQKNRLFFYRKSVWSKLQSIGI	13	EAEVR
spTRT	429	WLYNSFIIPILQSFFYITESSDLRNRTVYFRKDIWKLLCRPFI	12	ENNVR
Ea_p123	441	WIFEDLVVSLIRCFFYVTEQQKSYSKTYYYRKNIWDVIMKMSI	12	EKEVE
Sc_Est2	366	WLFRQLIPKIIQTFFYCTEISSTVT.IVYFRHDTWNKLITPFI	9	ENNVC

Telomerase RT Motifs (Fingers)

		MOTIF 1		MOTIF 2		MOTIF A		MOTIF B'	
TRT con		R iPKk		fR I		p lyF D cYD i		Y q GipQGs lS l	y
hTRT	11	SRLRFIPKPDG	0	LRPIV	69	PELYFVKVDVTGAYDTI	104	YVQCQGIPQGSILSTLLCSLCY	
spTRT	10	AVIRLLPKKNT	0	FRLIT	66	RKKYFVRIDIKSCYDRI	99	YLQKVGIPQGSILSSFLCHFYM	
Ea_p123	10	GKLRLIPKKTT	0	FRPIM	67	PKLFFATMDIEKCYDSV	117	YKQTKGIPQGLCVSSILSSFYY	
Sc_Est2	13	SKMRIIPKKSN	2	FRIIA	68	PELYFMKFDVKSCYDSI	85	YIREDGLFQGSSLSAPIVDLVY	
RT con		p hh h K		hR h		h hDh AF h GY		hPQG pP hh h	

Telomerase RT Motifs (Palm, Primer Grip)

		MOTIF C		MOTIF D		MOTIF E	
TRT con		lllrl DDfL it		g n K		w g s l	
hTRT	15	LLLRLVDDFLLVT	15	GVPEYGCVVNLRKTVV	24	WCGLLLDTRTL	192
spTRT	16	VLLRVVDDFLFIT	15	GFEKHNFSTSLEKTVI	22	FFGFSVNMRSL	176
Ea_p123	24	LLMRLTDDYLLIT	15	VSRENGFKFNMKKLQT	28	WIGISIDMKTL	174
Sc_Est2	18	LILKLADDFLIIS	15	GFQKYNAKANRDKILA	25	WKHSSTMNNFH	141
RT con		h Y DDhhh F		Gh h cK h		hLG h	

FIG. 11

```
181 GGACCCGGCGGCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGC
    CCTGGGCCGCCGAAAGGCGCGCGACCACCGGGTCACGGACCACACGCACGGGACCCTGCG

                                              NFkB_CS1
                                             GGGRQTYYQC
                                           NFkB-MHC-I.2
                                           TGGGCTTCCCC
                                          **************************
241 ACGGCCGCCCCCCGCCGCCCCCTCCTTCCGCCAGGTGGGCCTCCCCGGGGTCGGCGTCCG
    TGCCGGCGGGGGGCGGCGGGGGAGGAAGGCGGTCCACCCGGAGGGGCCCCAGCCGCAGGC

                              Intron1
    ************************************************************
301 GCTGGGGTTGAGGGCGGCCGGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGGCGACTC
    CGACCCCAACTCCCGCCGGCCCCCCTTGGTCGCTGTACGCCTCTCGTCGCGTCCGCTGAG

          NFkB_CS1
         GGGRQTYYQC
          NFkB_CS2
        RGGGRMTYYCC
           Topo_II_cleavage_site
             RNYNNCNNGYNGKTNYNY
    *****************>
361 AGGGCGCTTCCCCCGCAGGTGTCCTGCCTGAAGGAGCTGGTGGCCCGAGTGCTGCAGAGG
    TCCCGCGAAGGGGGCGTCCACAGGACGGACTTCCTCGACCACCGGGCTCACGACGTCTCC
```

FIG. 12

```
   1  AAAACCCCAA AACCCCAAAA CCCCTTTTAG AGCCCTGCAG TTGGAAATAT
  51  AACCTCAGTA TTAATAAGCT CAGATTTTAA ATATTAATTA CAAAACCTAA
 101  ATGGAGGTTG ATGTTGATAA TCAAGCTGAT AATCATGGCA TTCACTCAGC
 151  TCTTAAGACT TGTGAAGAAA TTAAAGAAGC TAAAACGTTG TACTCTTGGA
 201  TCCAGAAAGT TATTAGATGA AGAAATCAAT CTCAAAGTCA TTATAAAGAT
 251  TTAGAAGATA TTAAAATATT TGCGCAGACA AATATTGTTG CTACTCCACG
 301  AGACTATAAT GAAGAAGATT TTAAAGTTAT TGCAAGAAAA GAAGTATTTT
 351  CAACTGGACT AATGATCGAA CTTATTGACA AATGCTTAGT TGAACTTCTT
 401  TCATCAAGCG ATGTTTCAGA TAGACAAAAA CTTCAATGAT TTGGATTTCA
 451  ACTTAAGGGA AATCAATTAG CAAAGACCCA TTTATTAACA GCTCTTTCAA
 501  CTCAAAAGCA GTATTTCTTT CAAGACGAAT GGAACCAAGT TAGAGCAATG
 551  ATTGGAAATG AGCTCTTCCG ACATCTCTAC ACTAAATATT TAATATTCCA
 601  GCGAACTTCT GAAGGAACTC TTGTTCAATT TTGCGGGAAT AACGTTTTTG
 651  ATCATTTGAA AGTCAACGAT AAGTTTGACA AAAAGCAAAA AGGTGGAGCA
 701  GCAGACATGA ATGAACCTCG ATGTTGATCA ACCTGCAAAT ACAATGTCAA
 751  GAATGAGAAA GATCACTTTC TCAACAACAT CAACGTGCCG AATTGGAATA
 801  ATATGAAATC AAGAACCAGA ATATTTTATT GCACTCATTT TAATAGAAAT
 851  AACCAATTCT TCAAAAAGCA TGAGTTTGTG AGTAACAAAA ACAATATTTC
 901  AGCGATGGAC AGAGCTCAGA CGATATTCAC GAATATATTC AGATTTAATA
 951  GAATTAGAAA GAAGCTAAAA GATAAGGTTA TCGAAAAAAT TGCCTACATG
1001  CTTGAGAAAG TCAAAGATTT TAACTTCAAC TACTATTTAA CAAAATCTTG
1051  TCCTCTTCCA GAAAATTGGC GGGAACGGAA ACAAAAAATC GAAAACTTGA
1101  TAAATAAAAC TAGAGAAGAA AAGTCGAAGT ACTATGAAGA GCTGTTTAGC
1151  TACACAACTG ATAATAAATG CGTCACACAA TTTATTAATG AATTTTTCTA
1201  CAATATACTC CCCAAAGACT TTTTGACTGG AAGAAACCGT AAGAATTTTC
1251  AAAAGAAAGT TAAGAAATAT GTGGAACTAA ACAAGCATGA ACTCATTCAC
1301  AAAAACTTAT TGCTTGAGAA GATCAATACA AGAGAAATAT CATGGATGCA
1351  GGTTGAGACC TCTGCAAAGC ATTTTTATTA TTTTGATCAC GAAAACATCT
1401  ACGTCTTATG GAAATTGCTC CGATGGATAT TCGAGGATCT CGTCGTCTCG
1451  CTGATTAGAT GATTTTTCTA TGTCACCGAG CAACAGAAAA GTTACTCCAA
1501  AACCTATTAC TACAGAAAGA ATATTTGGGA CGTCATTATG AAAATGTCAA
1551  TCGCAGACTT AAAGAAGGAA ACGCTTGCTG AGGTCCAAGA AAAAGAGGTT
1601  GAAGAATGGA AAAAGTCGCT TGGATTTGCA CCTGGAAAAC TCAGACTAAT
1651  ACCGAAGAAA ACTACTTTCC GTCCAATTAT GACTTTCAAT AAGAAGATTG
1701  TAAATTCAGA CCGGAAGACT ACAAAATTAA CTACAAATAC GAAGTTATTG
1751  AACTCTCACT TAATGCTTAA GACATTGAAG AATAGAATGT TTAAAGATCC
1801  TTTTGGATTC GCTGTTTTTA ACTATGATGA TGTAATGAAA AAGTATGAGG
1851  AGTTTGTTTG CAAATGGAAG CAAGTTGGAC AACCAAAACT CTTCTTTGCA
1901  ACTATGGATA TCGAAAAGTG ATATGATAGT GTAAACAGAG AAAAACTATC
1951  AACATTCCTA AAAACTACTA AATTACTTTC TTCAGATTTC TGGATTATGA
2001  CTGCACAAAT TCTAAAGAGA AAGAATAACA TAGTTATCGA TTCGAAAAAC
2051  TTTAGAAAGA AAGAAATGAA AGATTATTTT AGACAGAAAT TCCAGAAGAT
2101  TGCACTTGAA GGAGGACAAT ATCCAACCTT ATTCAGTGTT CTTGAAAATG
2151  AACAAAATGA CTTAAATGCA AAGAAAACAT TAATTGTTGA AGCAAAGCAA
2201  AGAAATTATT TTAAGAAAGA TAACTTACTT CAACCAGTCA TTAATATTTG
2251  CCAATATAAT TACATTAACT TTAATGGGAA GTTTTATAAA CAAACAAAAG
2301  GAATTCCTCA AGGTCTTTGA GTTTCATCAA TTTTGTCATC ATTTTATTAT
2351  GCAACATTAG AGGAAAGCTC CTTAGGATTC CTTAGAGATG AATCAATGAA
```

FIG. 13A

```
2401  CCCTGAAAAT CCAAATGTTA ATCTTCTAAT GAGACTTACA GATGACTATC
2451  TTTTGATTAC AACTCAAGAG AATAATGCAG TATTGTTTAT TGAGAAACTT
2501  ATAAACGTAA GTCGTGAAAA TGGATTTAAA TTCAATATGA AGAAACTACA
2551  GACTAGTTTT CCATTAAGTC CAAGCAAATT TGCAAAATAC GGAATGGATA
2601  GTGTTGAGGA GCAAAATATT GTTCAAGATT ACTGCGATTG GATTGGCATC
2651  TCAATTGATA TGAAAACTCT TGCTTTAATG CCAAATATTA ACTTGAGAAT
2701  AGAAGGAATT CTGTGTACAC TCAATCTAAA CATGCAAACA AAGAAAGCAT
2751  CAATGTGGCT CAAGAAGAAA CTAAAGTCGT TTTTAATGAA TAACATTACC
2801  CATTATTTTA GAAAGACGAT TACAACCGAA GACTTTGCGA ATAAAACTCT
2851  CAACAAGTTA TTTATATCAG GCGGTTACAA ATACATGCAA TGAGCCAAAG
2901  AATACAAGGA CCACTTTAAG AAGAACTTAG CTATGAGCAG TATGATCGAC
2951  TTAGAGGTAT CTAAAATTAT ATACTCTGTA ACCAGAGCAT TCTTTAAATA
3001  CCTTGTGTGC AATATTAAGG ATACAATTTT TGGAGAGGAG CATTATCCAG
3051  ACTTTTTCCT TAGCACACTG AAGCACTTTA TTGAAATATT CAGCACAAAA
3101  AAGTACATTT TCAACAGAGT TTGCATGATC CTCAAGGCAA AAGAAGCAAA
3151  GCTAAAAAGT GACCAATGTC AATCTCTAAT TCAATATGAT GCATAGTCGA
3201  CTATTCTAAC TTATTTTGGA AAGTTAATTT TCAATTTTTG TCTTATATAC
3251  TGGGGTTTTG GGGTTTTGGG GTTTTGGGG
```

FIG. 13B

```
   1  MEVDVDNQAD NHGIHSALKT CEEIKEAKTL YSWIQKVIRC RNQSQSHYKD
  51  LEDIKIFAQT NIVATPRDYN EEDFKVIARK EVFSTGLMIE LIDKCLVELL
 101  SSSDVSDRQK LQCFGFQLKG NQLAKTHLLT ALSTQKQYFF QDEWNQVRAM
 151  IGNELFRHLY TKYLIFQRTS EGTLVQFCGN NVFDHLKVND KFDKKQKGGA
 201  ADMNEPRCCS TCKYNVKNEK DHFLNNINVP NWNNMKSRTR IFYCTHFNRN
 251  NQFFKKHEFV SNKNNISAMD RAQTIFTNIF RFNRIRKKLK DKVIEKIAYM
 301  LEKVKDFNFN YYLTKSCPLP ENWRERKQKI ENLINKTREE KSKYYEELFS
 351  YTTDNKCVTQ FINEFFYNIL PKDFLTGRNR KNFQKKVKKY VELNKHELIH
 401  KNLLLEKINT REISWMQVET SAKHFYYFDH ENIYVLWKLL RWIFEDLVVS
 451  LIRCFFYVTE QQKSYSKTYY YRKNIWDVIM KMSIADLKKE TLAEVQEKEV
 501  EEWKKSLGFA PGKLRLIPKK TTFRPIMTFN KKIVNSDRKT TKLTTNTKLL
 551  NSHLMLKTLK NRMFKDPFGF AVFNYDDVMK KYEEFVCKWK QVGQPKLFFA
 601  TMDIEKCYDS VNREKLSTFL KTTKLLSSDF WIMTAQILKR KNNIVIDSKN
 651  FRKKEMKDYF RQKFQKIALE GGQYPTLFSV LENEQNDLNA KKTLIVEAKQ
 701  RNYFKKDNLL QPVINICQYN YINFNGKFYK QTKGIPQGLC VSSILSSFYY
 751  ATLEESSLGF LRDESMNPEN PNVNLLMRLT DDYLLITTQE NNAVLFIEKL
 801  INVSRENGFK FNMKKLQTSF PLSPSKFAKY GMDSVEEQNI VQDYCDWIGI
 851  SIDMKTLALM PNINLRIEGI LCTLNLNMQT KKASMWLKKK LKSFLMNNIT
 901  HYFRKTITTE DFANKTLNKL FISGGYKYMQ CAKEYKDHFK KNLAMSSMID
 951  LEVSKIIYSV TRAFFKYLVC NIKDTIFGEE HYPDFFLSTL KHFIEIFSTK
1001  KYIFNRVCMI LKAKEAKLKS DQCQSLIQYD A
```

FIG. 14

```
   1 ggtaccgatttactttcctttcttcataagctaattgcttcctcgaacgctcctaaatctctggaaatatttttacaaga  80
  81 actcaataacaataccaagtcaaattccaatatgaaggtgttattagtgatcgataatatttctattttatcggtcgtta 160
 161 ccaagtataaggacaaaaagaacaacttccttccccctaaagacttttacttttattaatttacttttcaaatatatttcg 240
 241 ggttcgcttacttttaatcgtggtactgttttagctgctacttctagccaaccgcgtgtttctaccccgtcattggatat 320
 321 agctcttggagtagctcacagaaatccttacaaatcttctgatgagactatattagattcattacagtccgtgcatattc 400
 401 ttaacatggagccttacactttagatgagtcacgtcgcatgatggagtatttggtatcatccaacgtttgccttgaaaag 480
 481 gttgataattatttgcaaaatcatgtccttagtggtggtaatccgcgaaagtttttgatgcttgcacacgtctagcatg 560
 561 attgagatattcaaaaatttctatccactacaactcctttaacgcggttttatttttctattttctattctcatgttgtt 640
 641 ccaaatatgtatcatctcgtattaggctttttccgttttactcctggaatcgtacctttttcactattccccctaatga 720
 721 ataatctaaattagtttcgcttataattgatagtagtagaaagattggtgattctactcgtgtaatgttattagtttaaa 800
 801 gatactttgcaaaacatttattagctatcattatataaaaaaaatcctataattataaatattaatcaatatttgcggtc 880
 881 actatttatttaaaacgttatgatcagtaggacactttgcatatatatagttatgcttaatggttacttgtaacttgc   958

 959 ATG ACC GAA CAC CAT ACC CCC AAA AGC AGG ATT CTT CGC TTT CTA GAG AAT CAA TAT GTA  1018
   1 M   T   E   H   H   T   P   K   S   R   I   L   R   F   L   E   N   Q   Y   V    20

1019 TAC CTA TGT ACC TTA AAT GAT TAT GTA CAA CTT GTT TTG AGA GGG TCG CCG GCA AGC TCG  1078
  21 Y   L   C   T   L   N   D   Y   V   Q   L   V   L   R   G   S   P   A   S   S    40

1079 TAT AGC AAT ATA TGC GAA CGC TTG AGA AGC GAT GTA CAA ACG TCC TTT TCT ATT TTT CTT  1138
  41 Y   S   N   I   C   E   R   L   R   S   D   V   Q   T   S   F   S   I   F   L    60

1139 CAT TCG ACT GTA GTC GGC TTC GAC AGT AAG CCA GAT GAA GGT GTT CAA TTT TCT TCT CCA  1198
  61 H   S   T   V   V   G   F   D   S   K   P   D   E   G   V   Q   F   S   S   P    80

1199 AAA TGC TCA CAG TCA GAG gtatatatatttttgttttgatttttttctattcgggatagctaatatatgggcag 1272
  81 K   C   S   Q   S   E                                                           86

1273 CTA ATA GCG AAT GTT GTA AAA CAG ATG TTC GAT GAA AGT TTT GAG CGT CGA AGG AAT CTA  1332
  87 L   I   A   N   V   V   K   Q   M   F   D   E   S   F   E   R   R   R   N   L   106

1333 CTG ATG AAA GGG TTT TCC ATG gtaaggtattctaattgtgaaatatttacctgcaattactgtttcaaagaga 1405
 107 L   M   K   G   F   S   M                                                       113

1406 ttgtatttaaccgataaag AAT CAT GAA GAT TTT CGA GCC ATG CAT GTA AAC GGA GTA CAA AAT  1469
 114                     N   H   E   D   F   R   A   M   H   V   N   G   V   Q   N   128
```

FIG. 15A

```
1470 GAT CTC GTT TCT ACT TTT CCT AAT TAC CTT ATA TCT ATA CTT GAG TCA AAA AAT TGG CAA  1529
 129 D   L   V   S   T   F   P   N   Y   L   I   S   I   L   E   S   K   N   W   Q    148

1530 CTT TTG TTA GAA AT gtaaataccggttaagatgttgcgcactttgaacaagactgacaagtatag T ATC GGC 1601
 149 L   L   L   E   I                                                     I   G   155

1602 AGT GAT GCC ATG CAT TAC TTA TTA TCC AAA GGA AGT ATT TTT GAG GCT CTT CCA AAT GAC  1661
 156 S   D   A   M   H   Y   L   L   S   K   G   S   I   F   E   A   L   P   N   D    175

1662 AAT TAC CTT CAG ATT TCT GGC ATA CCA CTT TTT AAA AAT AAT GTG TTT GAG GAA ACT GTG  1721
 176 N   Y   L   Q   I   S   G   I   P   L   F   K   N   N   V   F   E   E   T   V    195

1722 TCA AAA AAA AGA AAG CGA ACC ATT GAA ACA TCC ATT ACT CAA AAT AAA AGC GCC CGC AAA  1781
 196 S   K   K   R   K   R   T   I   E   T   S   I   T   Q   N   K   S   A   R   K    215

1782 GAA GTT TCC TGG AAT AGC ATT TCA ATT AGT AGG TTT AGC ATT TTT TAC AGG TCA TCC TAT  1841
 216 E   V   S   W   N   S   I   S   I   S   R   F   S   I   F   Y   R   S   S   Y    235

1842 AAG AAG TTT AAG CAA G gtaactaatactgttatccttcataactaattttag AT CTA TAT TTT AAC    1907
 236 K   K   F   K   Q   D                                       L   Y   F   N      245

1908 TTA CAC TCT ATT TGT GAT CGG AAC ACA GTA CAC ATG TGG CTT CAA TGG ATT TTT CCA AGG  1967
 246 L   H   S   I   C   D   R   N   T   V   H   M   W   L   Q   W   I   F   P   R    265

1968 CAA TTT GGA CTT ATA AAC GCA TTT CAA GTG AAG CAA TTG CAC AAA GTG ATT CCA CTG GTA  2027
 266 Q   F   G   L   I   N   A   F   Q   V   K   Q   L   H   K   V   I   P   L   V    285

2028 TCA CAG AGT ACA GTT GTG CCC AAA CGT CTC CTA AAG GTA TAC CCT TTA ATT GAA CAA ACA  2087
 286 S   Q   S   T   V   V   P   K   R   L   L   K   V   Y   P   L   I   E   Q   T    305

2088 GCA AAG CGA CTC CAT CGT ATT TCT CTA TCA AAA GTT TAC AAC CAT TAT TGC CCA TAT ATT  2147
 306 A   K   R   L   H   R   I   S   L   S   K   V   Y   N   H   Y   C   P   Y   I    325

2148 GAC ACC CAC GAT GAT GAA AAA ATC CTT AGT TAT TCC TTA AAG CCG AAC CAG GTG TTT GCG  2207
 326 D   T   H   D   D   E   K   I   L   S   Y   S   L   K   P   N   Q   V   F   A    345

2208 TTT CTT CGA TCC ATT CTT GTT CGA GTG TTT CCT AAA TTA ATC TGG GGT AAC CAA AGG ATA  2267
 346 F   L   R   S   I   L   V   R   V   F   P   K   L   I   W   G   N   Q   R   I    365
```

FIG. 15B

```
2268 TTT GAG ATA ATA TTA AAA G gtattgtataaaatttattaccactaacgattttaccag AC CTC GAA ACT 2336
 366 F   E   I   I   L   K   D                                           L   E   T   375

2337 TTC TTG AAA TTA TCG AGA TAC GAG TCT TTT AGT TTA CAT TAT TTA ATG AGT AAC ATA AAG  2396
 376 F   L   K   L   S   R   Y   E   S   F   S   L   H   Y   L   M   S   N   I   K    395

2397 gtaatatgccaaatttttttaccattaattaacaatcag ATT TCA GAA ATT GAA TGG CTA GTC CTT GGA  2465
 396                                        I   S   E   I   E   W   L   V   L   G    405

2466 AAA AGG TCA AAT GCG AAA ATG TGC TTA AGT GAT TTT GAG AAA CGC AAG CAA ATA TTT GCG  2525
 406 K   R   S   N   A   K   M   C   L   S   D   F   E   K   R   K   Q   I   F   A    425

2526 GAA TTC ATC TAC TGG CTA TAC AAT TCG TTT ATA ATA CCT ATT TTA CAA TCT TTT TTT TAT  2585
 426 E   F   I   Y   W   L   Y   N   S   F   I   I   P   I   L   Q   S   F   F   Y    445

2586 ATC ACT GAA TCA AGT GAT TTA CGA AAT CGA ACT GTT TAT TTT AGA AAA GAT ATT TGG AAA  2645
 446 I   T   E   S   S   D   L   R   N   R   T   V   Y   F   R   K   D   I   W   K    465

2646 CTC TTG TGC CGA CCC TTT ATT ACA TCA ATG AAA ATG GAA GCG TTT GAA AAA ATA AAC GAG  2705
 466 L   L   C   R   P   F   I   T   S   M   K   M   E   A   F   E   K   I   N   E    485

2706 gtattttaaagtatttttttgcaaaaagctaatattttcag AAC AAT GTT AGG ATG GAT ACT CAG AAA ACT 2775
 486                                           N   N   V   R   M   D   T   Q   K   T   495

2776 ACT TTG CCT CCA GCA GTT ATT CGT CTA TTA CCT AAG AAG AAT ACC TTT CGT CTC ATT ACG  2835
 496 T   L   P   P   A   V   I   R   L   L   P   K   K   N   T   F   R   L   I   T    515

2836 AAT TTA AGA AAA AGA TTC TTA ATA AAG gtattaatttttggtcatcaatgtactttacttctaatctatta 2906
 516 N   L   R   K   R   F   L   I   K                                               524

2907 ttagcag ATG GGT TCA AAC AAA AAA ATG TTA GTC AGT ACG AAC CAA ACT TTA CGA CCT GTG  2967
 525         M   G   S   N   K   K   M   L   V   S   T   N   Q   T   L   R   P   V    542

2968 GCA TCG ATA CTG AAA CAT TTA ATC AAT GAA GAA AGT AGT GGT ATT CCA TTT AAC TTG GAG  3027
 543 A   S   I   L   K   H   L   I   N   E   E   S   S   G   I   P   F   N   L   E    562

3028 GTT TAC ATG AAG CTT CTT ACT TTT AAG AAG GAT CTT CTT AAG CAC CGA ATG TTT GG gtaat 3088
 563 V   Y   M   K   L   L   T   F   K   K   D   L   L   K   H   R   M   F   G       581
```

FIG. 15C

```
3089 tatataatgcgcgattcctcattattaattttgcag G CGT AAG AAG TAT TTT GTA CGG ATA GAT ATA   3155
 582                                        R   K   K   Y   F   V   R   I   D   I     591

3156 AAA TCC TGT TAT GAT CGA ATA AAG CAA GAT TTG ATG TTT CGG ATT GTT AAA AAG AAA CTC   3215
 592 K   S   C   Y   D   R   I   K   Q   D   L   M   F   R   I   V   K   K   K   L     611

3216 AAG GAT CCC GAA TTT GTA ATT CGA AAG TAT GCA ACC ATA CAT GCA ACA AGT GAC CGA GCT   3275
 612 K   D   P   E   F   V   I   R   K   Y   A   T   I   H   A   T   S   D   R   A     631

3276 ACA AAA AAC TTT GTT AGT GAG GCG TTT TCC TAT T gtaagtttatttttcattggaatttttaacaa   3343
 632 T   K   N   F   V   S   E   A   F   S   Y   F                                     643

3344 attctttttag TT GAT ATG GTG CCT TTT GAA AAA GTC GTG CAG TTA CTT TCT ATG AAA ACA   3405
 644               D   M   V   P   F   E   K   V   V   Q   L   L   S   M   K   T     659

3406 TCA GAT ACT TTG TTT GTT GAT TTT GTG GAT TAT TGG ACC AAA AGT TCT TCT GAA ATT TTT   3465
 660 S   D   T   L   F   V   D   F   V   D   Y   W   T   K   S   S   S   E   I   F     679

3466 AAA ATG CTC AAG GAA CAT CTC TCT GGA CAC ATT GTT AAG gtataccaattgttgaattgtaataaca   3532
 680 K   M   L   K   E   H   L   S   G   H   I   V   K                                 692

3533 ctaatgaaactag ATA GGA AAT TCT CAA TAC CTT CAA AAA GTT GGT ATC CCT CAG GGC TCA     3593
 693              I   G   N   S   Q   Y   L   Q   K   V   G   I   P   Q   G   S       708

3594 ATT CTG TCA TCT TTT TTG TGT CAT TTC TAT ATG GAA GAT TTG ATT GAT GAA TAC CTA TCG   3653
 709 I   L   S   S   F   L   C   H   F   Y   M   E   D   L   I   D   E   Y   L   S     728

3654 TTT ACG AAA AAG AAA GGA TCA GTG TTG TTA CGA GTA GTC GAC GAT TTC CTC TTT ATA ACA   3713
 729 F   T   K   K   K   G   S   V   L   L   R   V   V   D   D   F   L   F   I   T     748

3714 GTT AAT AAA AAG GAT GCA AAA AAA TTT TTG AAT TTA TCT TTA AGA G gtgagttgctgtcattcc   3777
 749 V   N   K   K   D   A   K   K   F   L   N   L   S   L   R   G                   764

3778 taagttctaaccgttgaag GA TTT GAG AAA CAC AAT TTT TCT ACG AGC CTG GAG AAA ACA GTA   3840
 765                       F   E   K   H   N   F   S   T   S   L   E   K   T   V     778

3841 ATA AAC TTT GAA AAT AGT AAT GGG ATA ATA AAC AAT ACT TTT TTT AAT GAA AGC AAG AAA   3900
 779 I   N   F   E   N   S   N   G   I   I   N   N   T   F   F   N   E   S   K   K     798
```

FIG. 15D

```
3901 AGA ATG CCA TTC TTC GGT TTC TCT GTG AAC ATG AGG TCT CTT GAT ACA TTG TTA GCA TGT  3960
 799 R   M   P   F   F   G   F   S   V   N   M   R   S   L   D   T   L   L   A   C    818

3961 CCT AAA ATT GAT GAA GCC TTA TTT AAC TCT ACA TCT GTA GAG CTG ACG AAA CAT ATG GGG  4020
 819 P   K   I   D   E   A   L   F   N   S   T   S   V   E   L   T   K   H   M   G    838

4021 AAA TCT TTT TTT TAC AAA ATT CTA AG gtatactgtgtaactgaataatagctgacaaataatcag A TCG 4089
 839 K   S   F   F   Y   K   I   L   R                                           S   848

4090 AGC CTT GCA TCC TTT GCA CAA GTA TTT ATT GAC ATT ACC CAC AAT TCA AAA TTC AAT TCT  4149
 849 S   L   A   S   F   A   Q   V   F   I   D   I   T   H   N   S   K   F   N   S    868

4150 TGC TGC AAT ATA TAT AGG CTA GGA TAC TCT ATG TGT ATG AGA GCA CAA GCA TAC TTA AAA  4209
 869 C   C   N   I   Y   R   L   G   Y   S   M   C   M   R   A   Q   A   Y   L   K    888

4210 AGG ATG AAG GAT ATA TTT ATT CCC CAA AGA ATG TTC ATA ACG G gtgagtacttattttaactaga 4274
 889 R   M   K   D   I   F   I   P   Q   R   M   F   I   T   D                       903

4275 aaagtcattaattaaccttag AT CTT TTG AAT GTT ATT GGA AGA AAA ATT TGG AAA AAG TTG GCC 4339
 904                         L   L   N   V   I   G   R   K   I   W   K   K   L   A   917

4340 GAA ATA TTA GGA TAT ACG AGT AGG CGT TTC TTG TCC TCT GCA GAA GTC AAA TG gtacgtgtc 4401
 918 E   I   L   G   Y   T   S   R   R   F   L   S   S   A   E   V   K   W           935

4402 ggtctcgagacttcagcaatattgacacatcag G CTT TTT TGT CTT GGA ATG AGA GAT GGT TTG AAA  4468
 936                                   L   F   C   L   G   M   R   D   G   L   K    946

4469 CCC TCT TTC AAA TAT CAT CCA TGC TTC GAA CAG CTA ATA TAC CAA TTT CAG TCA TTG ACT  4528
 947 P   S   F   K   Y   H   P   C   F   E   Q   L   I   Y   Q   F   Q   S   L   T    966

4529 GAT CTT ATC AAG CCG CTA AGA CCA GTT TTG CGA CAG GTG TTA TTT TTA CAT AGA AGA ATA  4588
 967 D   L   I   K   P   L   R   P   V   L   R   Q   V   L   F   L   H   R   R   I    986

4589 GCT GAT TAA tgtcattttcaatttattatatacatcctttattactggtgtcttaaacaatattattactaagtata 4665
 987 A   D   *                                                                       989
```

FIG. 15E

```
4666 gctgacccccaaagcaagcatactataggatttctagtaaagtaaaattaatctcgttattagttttgattgacttgtct 4745
4746 ttatccttatacttttaagaaagattgacagtggttgctgactactgcccacatgcccattaaacgggagtggttaaaca 4825
4826 ttaaaagtaatacatgaggctaatctcctttcatttagaataaggaaagtggttttctataatgaataatgcccgcacta 4905
4906 atgcaaaaagacgaagattatcttctaaacaaggggattaagcatatccgaaggaaaagagagtaatatacccagtgtt 4985
4986 gttgaagaaagcaaggataatttggaacaagcttctgcagatgacaggctaaattttggtgaccgaattttggtaaaagc 5065
5066 cccaggttatccatggtggccggccttgctactgagacgaaaagaaactaaggatagtttgaatactaatagctcattta 5145
5146 atgtcttatataaggttttgtttttcctgacttcaattttgcatgggtgaaaagaaatagtgttaagccattattggat 5225
5226 tccgaaatagccaaatttcttggttcctcaaagcggaagtctaaagaacttattgaagcttatgaggcttcaaaaactcc 5305
5306 tcctgatttaaaggaggaatcttccaccgatgaggaaatggatagcttatcagctgctgaggagaagcctaatttttgc 5385
5386 aaaaaagaaaatatcattgggagacatctcttgatgaatcagatgcggagagtatctccagcggatccttgatgtcaata 5465
5466 acttctatttctgaaatgtatggtcctactgtcgcttcgacttctcgtagctctacgcagttaagtgaccaaaggtacc  5544
```

FIG. 15F

```
   1 gcagcgctgc gtcctgctgc gcacgtggga agccctggcc ccggccaccc ccgcgatgcc
  61 gcgcgctccc cgctgccgag ccgtgcgctc cctgctgcgc agccactacc gcgaggtgct
 121 gccgctggcc acgttcgtgc ggcgcctggg gccccagggc tggcggctgg tgcagcgcgg
 181 ggacccggcg gctttccgcg cgctggtggc ccagtgcctg gtgtgcgtgc cctgggacgc
 241 acggccgccc cccgccgccc cctccttccg ccaggtgtcc tgcctgaagg agctggtggc
 301 ccgagtgctg cagaggctgt gcgagcgcgg cgcgaagaac gtgctggcct tcggcttcgc
 361 gctgctggac ggggcccgcg ggggccccc cgaggccttc accaccagcg tgcgcagcta
 421 cctgcccaac acggtgaccg acgcactgcg ggggagcggg gcgtgggggc tgctgctgcg
 481 ccgcgtgggc gacgacgtgc tggttcacct gctggcacgc tgcgcgctct ttgtgctggt
 541 ggctcccagc tgcgcctacc aggtgtgcgg gccgccgctg taccagctcg gcgctgccac
 601 tcaggcccgg cccccgccac acgctagtgg accccgaagg cgtctgggat gcgaacgggc
 661 ctggaaccat agcgtcaggg aggccggggt ccccctgggc ctgccagccc cgggtgcgag
 721 gaggcgcggg ggcagtgcca gccgaagtct gccgttgccc aagaggccca ggcgtggcgc
 781 tgcccctgag ccggagcgga cgcccgttgg gcaggggtcc tgggcccacc cgggcaggac
 841 gcgtggaccg agtgaccgtg gtttctgtgt ggtgtcacct gccagacccg ccgaagaagc
 901 cacctctttg gagggtgcgc tctctggcac gcgccactcc cacccatccg tgggccgcca
 961 gcaccacgcg ggccccccat ccacatcgcg gccaccacgt ccctgggaca cgccttgtcc
1021 cccggtgtac gccgagacca agcacttcct ctactcctca ggcgacaagg agcagctgcg
1081 gccctccttc ctactcagct ctctgaggcc cagcctgact ggcgctcgga ggctcgtgga
1141 gaccatcttt ctgggttcca ggccctggat gccagggact ccccgcaggt tgccccgcct
1201 gccccagcgc tactggcaaa tgcggccccc gtttctggag ctgcttggga accacgcgca
1261 gtgcccctac ggggtgctcc tcaagacgca ctgcccgctg cgagctgcgg tcaccccagc
1321 agccggtgtc tgtgcccggg agaagcccca gggctctgtg gcggcccccg aggaggagga
1381 cacagacccc cgtcgcctgg tgcagctgct ccgccagcac agcagccccc ggcaggtgta
1441 cggcttcgtg cgggcctgcc tgcgccggct ggtgccccca ggcctctggg gctccaggca
1501 caacgaacgc cgcttcctca ggaacaccaa gaagttcatc tccctgggga agcatgccaa
1561 gctctcgctg caggagctga cgtggaagat gagcgtgcgg gactgcgctt ggctgcgcag
1621 gagcccaggg gttggctgtg ttccggccgc agagcaccgt ctgcgtgagg agatcctggc
1681 caagttcctg cactggctga tgagtgtgta cgtcgtcgag ctgctcaggt ctttctttta
1741 tgtcacggag accacgtttc aaaagaacag gctctttttc taccggaaga gtgtctggag
1801 caagttgcaa agcattggaa tcagacagca cttgaagagg gtgcagctgc gggagctgtc
1861 ggaagcagag gtcaggcagc atcgggaagc caggcccgcc ctgctgacgt ccagactccg
1921 cttcatcccc aagcctgacg ggctgcggcc gattgtgaac atggactacg tcgtgggagc
1981 cagaacgttc cgcagagaaa agagggccga gcgtctcacc tcgagggtga aggcactgtt
2041 cagcgtgctc aactacgagc gggcgcggcg ccccggcctc ctgggcgcct ctgtgctggg
2101 cctggacgat atccacaggg cctggcgcac cttcgtgctg cgtgtgcggg cccaggaccc
2161 gccgcctgag ctgtactttg tcaaggtgga tgtgacgggc gcgtacgaca ccatccccca
2221 ggacaggctc acggaggtca tcgccagcat catcaaaccc cagaacacgt actgcgtgcg
2281 tcggtatgcc gtggtccaga aggccgccca tgggcacgtc cgcaaggcct tcaagagcca
2341 cgtctctacc ttgacagacc tccagccgta catgcgacag ttcgtggctc acctgcagga
2401 gaccagcccg ctgagggatg ccgtcgtcat cgagcagagc tcctccctga atgaggccag
2461 cagtggcctc ttcgacgtct tcctacgctt catgtgccac cacgccgtgc gcatcagggg
2521 caagtcctac gtccagtgcc aggggatccc gcagggctcc atcctctcca cgctgctctg
2581 cagcctgtgc tacggcgaca tggagaacaa gctgtttgcg gggattcggc gggacgggct
2641 gctcctgcgt ttggtggatg atttcttgtt ggtgacacct cacctcaccc acgcgaaaac
2701 cttcctcagg accctggtcc gaggtgtccc tgagtatggc tgcgtggtga acttgcggaa
2761 gacagtggtg aacttccctg tagaagacga ggccctgggt ggcacggctt ttgttcagat
2821 gccggcccac ggcctattcc cctggtgcgg cctgctgctg gatacccgga ccctggaggt
2881 gcagagcgac tactccagct atgcccggac ctccatcaga gccagtctca ccttcaaccg
2941 cggcttcaag gctgggagga acatgcgtcg caaactcttt ggggtcttgc ggctgaagtg
3001 tcacagcctg tttctggatt tgcaggtgaa cagcctccag acggtgtgca ccaacatcta
3061 caagatcctc ctgctgcagg cgtacaggtt tcacgcatgt gtgctgcagc tcccatttca
3121 tcagcaagtt tggaagaacc ccacattttt cctgcgcgtc atctctgaca cggcctccct
3181 ctgctactcc atcctgaaag ccaagaacgc agggatgtcg ctgggggcca agggcgccgc
3241 cggccctctg ccctccgagg ccgtgcagtg gctgtgccac caagcattcc tgctcaagct
3301 gactcgacac cgtgtcacct acgtgccact cctggggtca ctcaggacag cccagacgca
3361 gctgagtcgg aagctcccgg ggacgacgct gactgccctg gaggccgcag ccaacccggc
3421 actgccctca gacttcaaga ccatcctgga ctgatggcca cccgcccaca gccaggccga
3481 gagcagacac cagcagccct gtcacgccgg gctctacgtc ccagggaggg aggggcggcc
3541 cacacccagg cccgcaccgc tgggagtctg aggcctgagt gagtgtttgg ccgaggcctg
3601 catgtccggc tgaaggctga gtgtccggct gaggcctgag cgagtgtcca gccaagggct
3661 gagtgtccag cacacctgcc gtcttcactt ccccacaggc tggcgctcgg ctccacccca
3721 gggccagctt ttcctcacca ggagcccggc ttccactccc cacataggaa tagtccatcc
3781 ccagattcgc cattgttcac ccctcgccct gccctccttt gccttccacc cccaccatcc
3841 aggtggagac cctgagaagg accctgggag ctctgggaat ttggagtgac caaaggtgtg
3901 ccctgtacac aggcgaggac cctgcacctg gatgggggtc cctgtgggtc aaattggggg
3961 gaggtgctgt gggagtaaaa tactgaatat atgagttttt cagttttgaa aaaaa
```

FIG. 16

MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRL
CERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALR
GSGAWGLLLRRVGDDVLVHLLARCALFVLVAPSCAYQVCGPPLY
QLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPAPG
ARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRG
PSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPP
STSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRP
SLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLEL
LGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEE
EDTDPRRLVQLLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNE
RRFLRNTKKFISLGKHAKLSLQELTWKMSVRDCAWLRRSPGVGC
VPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTETTFQKNR
LFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPAL
LTSRLRFIPKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKA
LFSVLNYERARRPGLLGASVLGLDDIHRAWRTFVLRVRAQDPPP
ELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYCVRRYAVVQ
KAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVI
EQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSI
LSTLLCSLCYGDMENKLFAGIRRDGLLLRLVDDFLLVTPHLTHA
KTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEALGGTAFVQMPA
HGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGR
NMRRKLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRF
HACVLQLPFHQQVWKNPTFFLRVISDTASLCYSILKAKNAGMSL
GAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGSLRTAQ
TQLSRKLPGTTLTALEAAANPALPSDFKTILD

FIG. 17

GGCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTT
TTATGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTG
GAGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCT
GTCGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACT
CCGCTTCATCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGG
AGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACT
GTTCAGCGTGCTCAACTACGAGCGGGCGCGGCGCCCCGGCCTCCTGGGCGCCTCTGTGCT
GGGCCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCCAGGA
CCCGCCGCCTGAGCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCC
CCAGGACAGGCTCACGGAGGTCATCGCCAGCATCATCAAACCCCAGAACACGTACTGCGT
GCGTCGGTATGCCGTGGTCCAGAAGGCCGCCCATGGGCACGTCCGCAAGGCCTTCAAGAG
CCACGTCCTACGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCT
GCAGCCTGTGCTACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTCGGCGGGACGGGC
TGCTCCTGCGTTTGGTGGATGATTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAA
CCTTCCTCAGGACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGA
AGACAGTGGTGAACTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGA
TGCCGGCCCACGGCCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGG
TGCAGAGCGACTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACC
GCGGCTTCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGT
GTCACAGCCTGTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCT
ACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTC
ATCAGCAAGTTTGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGGCCTCCC
TCTGCTACTCCATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGCGCCG
CCGGCC7TCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGC
TGACTCGACACCGTGTCACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGC
AGCTGAGTCGGAAGCTCCCGGGGACGACGCTGACTGCCCTGGAGGCCGCAGCCAACCCGG
CACTGCCCTCAGACTTCAAGACCATCCTGGACTGATGGCCACCCGCCCACAGCCAGGCCG
AGAGCAGACACCAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGC
CCACACCCAGGCCTGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCT
GCATGTCCGGCTGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGC
TGAGTGTCCAGCACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCC
AGGGCCAGCTTTTCCTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATC
CCCAGATTCGCCATTGTTCACCCCTCGCCCTGCCCTCCTTTGCCTTCCACCCCCACCATC
CAGGTGGAGACCCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGT
GCCCTGTACACAGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGG
GGAGGTGCTGTGGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTG0AAAAAAAAAA
AAAAAAAAAAAAAAAA

FIG. 18

MetSerValTyrValValGluLeuLeuArgSerPhePhe
TyrValThrGluThrThrPheGlnLysAsnArgLeuPhe
PheTyrArgLysSerValTrpSerLysLeuGlnSerIle
GlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu
LeuSerGluAlaGluValArgGlnHisArgGluAlaArg
ProAlaLeuLeuThrSerArgLeuArgPheIleProLys
ProAspGlyLeuArgProIleValAsnMetAspTyrVal
ValGlyAlaArgThrPheArgArgGluLysArgAlaGlu
ArgLeuThrSerArgValLysAlaLeuPheSerValLeu
AsnTyrGluArgAlaArgArgProGlyLeuLeuGlyAla
SerValLeuGlyLeuAspAspIleHisArgAlaTrpArg
ThrPheValLeuArgValArgAlaGlnAspProProPro
GluLeuTyrPheValLysValAspValThrGlyAlaTyr
AspThrIleProGlnAspArgLeuThrGluValIleAla
SerIleIleLysProGlnAsnThrTyrCysValArgArg
TyrAlaValValGlnLysAlaAlaHisGlyHisValArg
LysAlaPheLysSerHisValLeuArgProValProGly
AspProAlaGlyLeuHisProLeuHisAlaAlaLeuGln
ProValLeuArgArgHisGlyGluGlnAlaValCysGly
AspSerAlaGlyArgAlaAlaProAlaPheGlyGly

FIG. 19

```
                                                                    1
                                                                   met
GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCG ATG

                                        10
pro arg ala pro arg cys arg ala val arg ser leu leu arg ser
CCG CGC GCT CCC CGC TGC CGA GCC GTG CGC TCC CTG CTG CGC AGC

             20                                             30
his tyr arg glu val leu pro leu ala thr phe val arg arg leu
CAC TAC CGC GAG GTG CTG CCG CTG GCC ACG TTC GTG CGG CGC CTG

                                        40
gly pro gln gly trp arg leu val gln arg gly asp pro ala ala
GGG CCC CAG GGC TGG CGG CTG GTG CAG CGC GGG GAC CCG GCG GCT

             50                                             60
phe arg ala leu val ala gln cys leu val cys val pro trp asp
TTC CGC GCG CTG GTG GCC CAG TGC CTG GTG TGC GTG CCC TGG GAC

                                        70
ala arg pro pro pro ala ala pro ser phe arg gln val ser cys
GCA CGG CCG CCC CCC GCC GCC CCC TCC TTC CGC CAG GTG TCC TGC

             80                                             90
leu lys glu leu val ala arg val leu gln arg leu cys glu arg
CTG AAG GAG CTG GTG GCC CGA GTG CTG CAG AGG CTG TGC GAG CGC

                                       100
gly ala lys asn val leu ala phe gly phe ala leu leu asp gly
GGC GCG AAG AAC GTG CTG GCC TTC GGC TTC GCG CTG CTG GAC GGG

            110                                            120
ala arg gly gly pro pro glu ala phe thr thr ser val arg ser
GCC CGC GGG GGC CCC CCC GAG GCC TTC ACC ACC AGC GTG CGC AGC
```

FIG. 20A

```
                                        130
tyr leu pro asn thr val thr asp ala leu arg gly ser gly ala
TAC CTG CCC AAC ACG GTG ACC GAC GCA CTG CGG GGG AGC GGG GCG

            140                                         150
trp gly leu leu leu arg arg val gly asp asp val leu val his
TGG GGG CTG CTG CTG CGC CGC GTG GGC GAC GAC GTG CTG GTT CAC

                                        160
leu leu ala arg cys ala leu phe val leu val ala pro ser cys
CTG CTG GCA CGC TGC GCG CTC TTT GTG CTG GTG GCT CCC AGC TGC

            170                                         180
ala tyr gln val cys gly pro pro leu tyr gln leu gly ala ala
GCC TAC CAG GTG TGC GGG CCG CCG CTG TAC CAG CTC GGC GCT GCC

                                        190
thr gln ala arg pro pro pro his ala ser gly pro arg arg arg
ACT CAG GCC CGG CCC CCG CCA CAC GCT AGT GGA CCC CGA AGG CGT

            200                                         210
leu gly cys glu arg ala trp asn his ser val arg glu ala gly
CTG GGA TGC GAA CGG GCC TGG AAC CAT AGC GTC AGG GAG GCC GGG

                                        220
val pro leu gly leu pro ala pro gly ala arg arg arg gly gly
GTC CCC CTG GGC CTG CCA GCC CCG GGT GCG AGG AGG CGC GGG GGC

            230                                         240
ser ala ser arg ser leu pro leu pro lys arg pro arg arg gly
AGT GCC AGC CGA AGT CTG CCG TTG CCC AAG AGG CCC AGG CGT GGC

                                        250
ala ala pro glu pro glu arg thr pro val gly gln gly ser trp
GCT GCC CCT GAG CCG GAG CGG ACG CCC GTT GGG CAG GGG TCC TGG

            260                                         270
ala his pro gly arg thr arg gly pro ser asp arg gly phe cys
GCC CAC CCG GGC AGG ACG CGT GGA CCG AGT GAC CGT GGT TTC TGT


                                        280
val val ser pro ala arg pro ala glu glu ala thr ser leu glu
GTG GTG TCA CCT GCC AGA CCC GCC GAA GAA GCC ACC TCT TTG GAG

            290                                         300
gly ala leu ser gly thr arg his ser his pro ser val gly arg
GGT GCG CTC TCT GGC ACG CGC CAC TCC CAC CCA TCC GTG GGC CGC

                                        310
gln his his ala gly pro pro ser thr ser arg pro pro arg pro
CAG CAC CAC GCG GGC CCC CCA TCC ACA TCG CGG CCA CCA CGT CCC

            320                                         330
trp asp thr pro cys pro pro val tyr ala glu thr lys his phe
TGG GAC ACG CCT TGT CCC CCG GTG TAC GCC GAG ACC AAG CAC TTC
```

FIG. 20B

```
                                    340
leu tyr ser ser gly asp lys glu gln leu arg pro ser phe leu
CTC TAC TCC TCA GGC GAC AAG GAG CAG CTG CGG CCC TCC TTC CTA

            350                                     360
leu ser ser leu arg pro ser leu thr gly ala arg arg leu val
CTC AGC TCT CTG AGG CCC AGC CTG ACT GGC GCT CGG AGG CTC GTG

                                    370
glu thr ile phe leu gly ser arg pro trp met pro gly thr pro
GAG ACC ATC TTT CTG GGT TCC AGG CCC TGG ATG CCA GGG ACT CCC

            380                                     390
arg arg leu pro arg leu pro gln arg tyr trp gln met arg pro
CGC AGG TTG CCC CGC CTG CCC CAG CGC TAC TGG CAA ATG CGG CCC

                                    400
leu phe leu glu leu leu gly asn his ala gln cys pro tyr gly
CTG TTT CTG GAG CTG CTT GGG AAC CAC GCG CAG TGC CCC TAC GGG

            410                                     420
val leu leu lys thr his cys pro leu arg ala ala val thr pro
GTG CTC CTC AAG ACG CAC TGC CCG CTG CGA GCT GCG GTC ACC CCA

                                    430
ala ala gly val cys ala arg glu lys pro gln gly ser val ala
GCA GCC GGT GTC TGT GCC CGG GAG AAG CCC CAG GGC TCT GTG GCG

            440                                     450
ala pro glu glu glu asp thr asp pro arg arg leu val gln leu
GCC CCC GAG GAG GAG GAC ACA GAC CCC CGT CGC CTG GTG CAG CTG

                                    460
leu arg gln his ser ser pro trp gln val tyr gly phe val arg
CTC CGC CAG CAC AGC AGC CCC TGG CAG GTG TAC GGC TTC GTG CGG

            470                                     480
ala cys leu arg arg leu val pro pro gly leu trp gly ser arg
GCC TGC CTG CGC CGG CTG GTG CCC CCA GGC CTC TGG GGC TCC AGG

                                    490
his asn glu arg arg phe leu arg asn thr lys lys phe ile ser
CAC AAC GAA CGC CGC TTC CTC AGG AAC ACC AAG AAG TTC ATC TCC

            500                                     510
leu gly lys his ala lys leu ser leu gln glu leu thr trp lys
CTG GGG AAG CAT GCC AAG CTC TCG CTG CAG GAG CTG ACG TGG AAG

                                    520
met ser val arg asp cys ala trp leu arg arg ser pro gly val
ATG AGC GTG CGG GAC TGC GCT TGG CTG CGC AGG AGC CCA GGG GTT

            530                                     540
gly cys val pro ala ala glu his arg leu arg glu glu ile leu
GGC TGT GTT CCG GCC GCA GAG CAC CGT CTG CGT GAG GAG ATC CTG
```

FIG. 20C

```
                                    550
ala lys phe leu his trp leu met ser val tyr val val glu leu
GCC AAG TTC CTG CAC TGG CTG ATG AGT GTG TAC GTC GTC GAG CTG

            560                                         570
leu arg ser phe phe tyr val thr glu thr thr phe gln lys asn
CTC AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC

                                    580
arg leu phe phe tyr arg lys ser val trp ser lys leu gln ser
AGG CTC TTT TTC TAC CGG AAG AGT GTC TGG AGC AAG TTG CAA AGC
            590                                         600
ile gly ile arg gln his leu lys arg val gln leu arg glu leu
ATT GGA ATC AGA CAG CAC TTG AAG AGG GTG CAG CTG CGG GAG CTG

                                    610
ser glu ala glu val arg gln his arg glu ala arg pro ala leu
TCG GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG

            620                                         630
leu thr ser arg leu arg phe ile pro lys pro asp gly leu arg
CTG ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG

                                    640
pro ile val asn met asp tyr val val gly ala arg thr phe arg
CCG ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC

            650                                         660
arg glu lys arg ala glu arg leu thr ser arg val lys ala leu
AGA GAA AAG AGG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG

                                    670
phe ser val leu asn tyr glu arg ala arg arg pro gly leu leu
TTC AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG

            680                                         690
gly ala ser val leu gly leu asp asp ile his arg ala trp arg
GGC GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC

                                    700
thr phe val leu arg val arg ala gln asp pro pro pro glu leu
ACC TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG

            710                                         720
tyr phe val lys val asp val thr gly ala tyr asp thr ile pro
TAC TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC

                                    730
gln asp arg leu thr glu val ile ala ser ile ile lys pro gln
CAG GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG

            740                                         750
asn thr tyr cys val arg arg tyr ala val val gln lys ala ala
AAC ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC
```

FIG. 20D

```
                                        760
his gly his val arg lys ala phe lys ser his val leu arg pro
CAT GGG CAC GTC CGC AAG GCC TTC AAG AGC CAC GTC CTA CGT CCA

            770                                         780
val pro gly asp pro ala gly leu his pro leu his ala ala leu
GTG CCA GGG GAT CCC GCA GGG CTC CAT CCT CTC CAC GCT GCT CTG

                                    790
gln pro val leu arg arg his gly glu gln ala val cys gly asp
CAG CCT GTG CTA CGG CGA CAT GGA GAA CAA GCT GTT TGC GGG GAT

            800                             807
ser ala gly arg ala ala pro ala phe gly gly OP
TCG GCG GGA CGG GCT GCT CCT GCG TTT GGT GGA TGA TTTCTTGTTGGT

GACACCTCACCTCACCCACGCGAAAACCTTCCTCAGGACCCTGGTCCGAGGTGTCCCTGA

GTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGGTGAACTTCCCTGTAGAAGACGAGGC

CCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCCACGGCCTATTCCCCTGGTGCGGCCT

GCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCGACTACTCCAGCTATGCCCGGACCTC

CATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAAGGCTGGGAGGAACATGCGTCGCAA

ACTCTTTGGGGTCTTGCGGCTGAAGTGTCACAGCCTGTTTCTGGATTTGCAGGTGAACAG

CCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCA

CGCATGTGTGCTGCAGCTCCCATTTCATCAGCAAGTTTGGAAGAACCCCACATTTTTCCT

GCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAGG

GATGTCGCTGGGGGCCAAGGGCGCCGCCGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCT

GTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACACCGTGTCACCTACGTGCCACTCCT

GGGGTCACTCAGGACAGCCCAGACGCAGCTGAGTCGGAAGCTCCCGGGGACGACGCTGAC

TGCCCTGGAGGCCGCAGCCAACCCGGCACTGCCCTCAGACTTCAAGACCATCCTGGACTG

ATGGCCACCCGCCCACAGCCAGGCCGAGAGCAGACACCAGCAGCCCTGTCACGCCGGGCT

CTACGTCCCAGGGAGGGAGGGGCGGCCCACACCCAGGCCCGCACCGCTGGGAGTCTGAGG

CCTGAGTGAGTGTTTGGCCGAGGCCTGCATGTCCGGCTGAAGGCTGAGTGTCCGGCTGAG

GCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTCCAGCACACCTGCCGTCTTCACTTCCC

CACAGGCTGGCGCTCGGCTCCACCCCAGGGCCAGCTTTTCCTCACCAGGAGCCCGGCTTC

CACTCCCCACATAGGAATAGTCCATCCCCAGATTCGCCATTGTTCACCCCTCGCCCTGCC

CTCCTTTGCCTTCCACCCCCACCATCCAGGTGGAGACCCTGAGAAGGACCCTGGGAGCTC

TGGGAATTTGGAGTGACCAAAGGTGTGCCCTGTACACAGGCGAGGACCCTGCACCTGGAT

GGGGGTCCCTGTGGGTCAAATTGGGGGGAGGTGCTGTGGGAGTAAAATACTGAATATATG

AGTTTTTCAGTTTTGAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

FIG. 20E

```
3601 ATCGATTGGGCCCGAGATCTCGCGCGCGAGGCCTGCCATGGGACCCACTGCAGGGGCAGC
     TAGCTAACCCGGGCTCTAGAGCGCGCGCTCCGGACGGTACCCTGGGTGACGTCCCCGTCG
                    ^                       ^
                  3615                    3636
                  BGL2                    NCO1

3661 TGGGANGCTGCAGGCTTCAGGTCCCAGTGGGGTTGCCATCTGCCAGTAGAAACCTGATGT
     ACCCTNCGACGTCCGAAGTCCAGGGTCACCCCAACGGTAGACGGTCATCTTTGGACTACA

3721 AGAATCAGGGCGCGAGTGTGGACACTGTCCTGAATCTCAATGTCTCAGTGTGTGCTGAAA
     TCTTAGTCCCGCGCTCACACCTGTGACAGGACTTAGAGTTACAGAGTCACACACGACTTT

3781 CATGTAGAAATTAAAGTCCATCCCTCCTACTCTACTGGGATTGAGCCCCTTCCCTATCCC
     GTACATCTTTAATTTCAGGTAGGGAGGATGAGATGACCCTAACTCGGGGAAGGGATAGGG

3841 CCCCCAGGGGCAGAGGAGTTCCTCTCACTCCTGTGGAGGAAGGAATGATACTTTGTTATT
     GGGGGTCCCCGTCTCCTCAAGGAGAGTGAGGACACCTCCTTCCTTACTATGAAACAATAA

                                                                *****
3901 TTTCACTGCTGGTACTGAATCCACTGTTTCATTTGTTGGTTTGTTTGTTTTGTTTTGAGA
     AAAGTGACGACCATGACTTAGGTGACAAAGTAAACAACCAAACAAACAAAACAAAACTCT

     ************************************************************
3961 AGCGGTTTCACTCTTGTTGCTCAGGCTGGANGGAGTGCAATGGCGCGATCTTGGCTTACT
     TCGCCAAAGTGAGAACAACGAGTCCGACCTNCCTCACGTTACCGCGCTAGAACCGAATGA

           ALU
     ************************************************************
4021 GCAGCCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCTTCCGCCTCCCATTTGGCTGGGA
     CGTCGGAGACGGAGGGTCCAAGTTCACTAAGAGGACGAAGGCGGAGGGTAAACCGACCCT

     **************************          ************
4081 TTACAGGCACCCGCCACCATGCCCAGCTAATTTTTTGTATTTTTAGTANANACNGGGGTG
     AATGTCCGTGGGCGGTGGTACGGGTCGATTAAAAAACATAAAAATCATNTNTGNCCCCAC

                                                                A
       ============================================================
4141 GGGGTGGGGTTCACATGTTGGCCAAGCTGGTCTCGAACTTCTGAACTCAGATGATCCANC
     CCCCACCCCAAGTGTACAACCGGTTCGACCAGAGCTTGAAGACTTGAGTCTACTAGGTNG

     LU
     ============================================================
4201 TGCCTCTGCCTCCTAAAATTGCTGGGATTACAGGTGTNANCCACCATGCCCAACTCAAAA
     ACGGAGACGGAGGATTTTAACGACCCTAATGTCCACANTNGGTGGTACGGGTTGAGTTTT

4261 TTTACTCTGTTTANAAACATCTGGGTCTAAGGTAGGAANCTCACCCCACTCAATTTTTGT
     AAATGAGACAAATNTTTGTAGACCCAGATTCCATCCTTNGAGTGGGGTGAGTTAAAAACA
```

FIG. 21A

```
4321 GGTGTTTTTAAGCCAATNANAAAATTTTTTNATGTTGTTTNNNNNNNNNNNNNNNNNNNN
     CCACAAAAATTCGGTTANTNTTTTAAAAAANTACAACAAANNNNNNNNNNNNNNNNNNNN

4381 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4441 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4501 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4561 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4681 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4741 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4801 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4861 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4921 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

4981 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGGTGNNNGAGGG
     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCCACNNNCTCCC

5041 NGCCANGRAGGGGGCCAGGTTCCAANTTCCCAACCKTTTTWGGARGGACNGCCCCCAGGG
     NCGGTNCYTCCCCCGGTCCAAGGTTNAAGGGTTGGMAAAAWCCTYCCTGNCGGGGGTCCC

5101 GGGGATRAACAGANTNGGGGGKGGTWGGGTTNAKGGTGGGAACNCCTTNGCGCCTGGAG
     CCCCTAYTTGTCTNANCCCCCMCCAWCCCAANTMCCACCCTTGNGGAANCGSCGGACCTC

5161 AACGTGCAAAGAGGAAATGAAGGGCCTGKGTCAAGGAGCCCAAGTNGGCGGGGRAGTTTG
     TTGCACGTTTCTCCTTTACTTCCCGGACMCAGTTCCTCGGGTTCANCCGCCCCYTCAAAC

5221 CAGGGAGGCACTCCGGGGAGGTCCSGCGTGCCCGTCCAAGGGAGCAATGCGTCCTTCGGG
     GTCCCTCCGTGAGGCCCCTCCAGGSCGCACGGGCAGGTTCCCTCGTTACGCAGGAAGCCC

5281 TTCGTCCCCAWGCCGCGTCTACGCGCCTYCCGTCCTCCCCTTCACGTTCCGGCATTCGTG
     AAGCAGGGGTWCGGCGCAGATGCGCGGARGGCAGGAGGGGAAGTGCAAGGCCGTAAGCAC

5341 GTGCCCGGAGCCCGACGCCCCGCGTCCGGACCTGGAGGCAGCCCTGGGTCTCCGGATCAG
     CACGGGCCTCGGGCTGCGGGGCGCAGGCCTGGACCTCCGTCGGGACCCAGAGGCCTAGTC

5401 GCCAGCGGCCAAAGGGTCGCCGCACGCACCTGTTCCCAGGGCCTCCACATCATGGCCCCT
     CGGTCGCCGGTTTCCCAGCGGCGTGCGTGGACAAGGGTCCCGGAGGTGTAGTACCGGGGA
```

FIG. 21B

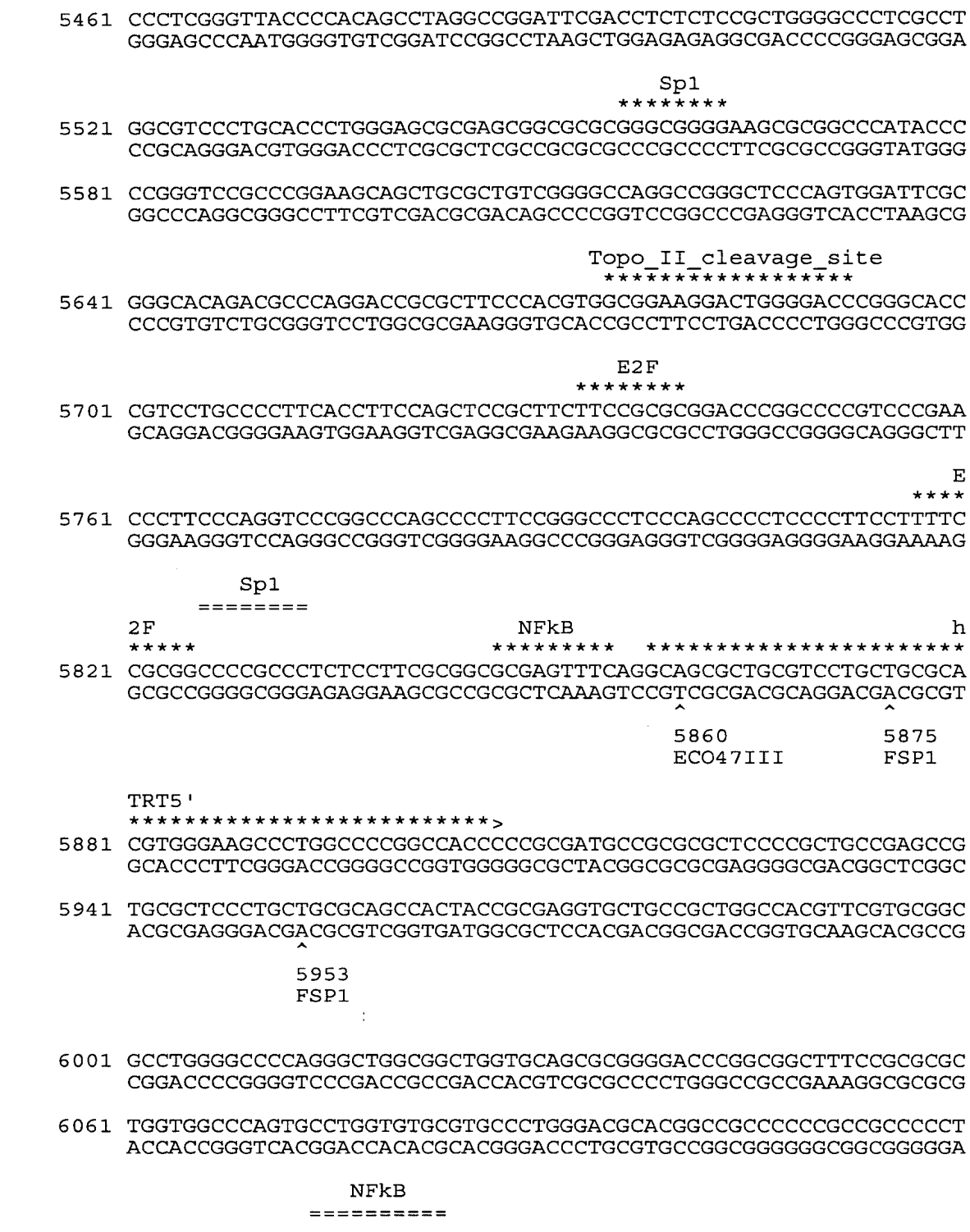
5461 CCCTCGGGTTACCCCACAGCCTAGGCCGGATTCGACCTCTCTCCGCTGGGGCCCTCGCCT
GGGAGCCCAATGGGGTGTCGGATCCGGCCTAAGCTGGAGAGAGGCGACCCCGGGAGCGGA
Sp1
****
5521 GGCGTCCCTGCACCCTGGGAGCGCGAGCGGCGCGCGGGCGGGGAAGCGCGGCCCATACCC
CCGCAGGGACGTGGGACCCTCGCGCTCGCCGCGCGCCCGCCCCTTCGCGCCGGGTATGGG
5581 CCGGGTCCGCCCGGAAGCAGCTGCGCTGTCGGGGCCAGGCCGGGCTCCCAGTGGATTCGC
GGCCCAGGCGGGCCTTCGTCGACGCGACAGCCCCGGTCCGGCCCGAGGGTCACCTAAGCG
Topo_II_cleavage_site
****
5641 GGGCACAGACGCCCAGGACCGCGCTTCCCACGTGGCGGAAGGACTGGGGACCCGGGCACC
CCCGTGTCTGCGGGTCCTGGCGCGAAGGGTGCACCGCCTTCCTGACCCCTGGGCCCGTGG
E2F
****
5701 CGTCCTGCCCCTTCACCTTCCAGCTCCGCTTCTTCCGCGCGGACCCGGCCCCGTCCCGAA
GCAGGACGGGGAAGTGGAAGGTCGAGGCGAAGAAGGCGCGCCTGGGCCGGGGCAGGGCTT
E
****
5761 CCCTTCCCAGGTCCCGGCCCAGCCCCTTCCGGGCCCTCCCAGCCCCTCCCCTTCCTTTTC
GGGAAGGGTCCAGGGCCGGGTCGGGGAAGGCCCGGGAGGGTCGGGGAGGGGAAGGAAAAG
Sp1
========
2F NFkB h
***** ********* **********
5821 CGCGGCCCCGCCCTCTCCTTCGCGGCGCGAGTTTCAGGCAGCGCTGCGTCCTGCTGCGCA
GCGCCGGGGCGGGAGAGGAAGCGCCGCGCTCAAAGTCCGTCGCGACGCAGGACGACGCGT
^ ^
5860 5875
ECO47III FSP1
TRT5'
**************************>
5881 CGTGGGAAGCCCTGGCCCCGGCCACCCCCGCGATGCCGCGCGCTCCCCGCTGCCGAGCCG
GCACCCTTCGGGACCGGGGCCGGTGGGGGCGCTACGGCGCGCGAGGGGCGACGGCTCGGC
5941 TGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGTTCGTGCGGC
ACGCGAGGGACGACGCGTCGGTGATGGCGCTCCACGACGGCGACCGGTGCAAGCACGCCG
^
5953
FSP1
6001 GCCTGGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTCCGCGCGC
CGGACCCCGGGGTCCCGACCGCCGACCACGTCGCGCCCCTGGGCCGCCGAAAGGCGCGCG
6061 TGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCCCCCCGCCGCCCCCT
ACCACCGGGTCACGGACCACACGCACGGGACCCTGCGTGCCGGCGGGGGGCGGCGGGGGA
NFkB
==========

FIG. 21C

```
                  **************************************************
6121 CCTTCCGCCAGGTGGGCCTCCCCGGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGG
     GGAAGGCGGTCCACCCGGAGGGGCCCCAGCCGCAGGCCGACCCCAACTCCCGCCGGCCCC

                                                    Topo_II_cleavage_s
                                                     ::::::::::::::::
                                                  NFkB
                                                ===========
     Intron1
     **************************************************>
6181 GGAACCAGCGACATGCGGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCCCGCAGGTGTC
     CCTTGGTCGCTGTACGCCTCTCGTCGCGTCCGCTGAGTCCCGCGAAGGGGGCGTCCACAG

     ite
     :



6241 CTGCCTGAAGGAGCTGGTGGCCCGAGTGCTGCAGAGGCTGTGCGAGCGCGGCGCGAAGAA
     GACGGACTTCCTCGACCACCGGGCTCACGACGTCTCCGACACGCTCGCGCCGCGCTTCTT

6301 CGTGCTGGCCTTCGGCTTCGCGCTGCTGGACGGGGCCCGCGGGGGCCCCCCCCGAGGCCTT
     GCACGACCGGAAGCCGAAGCGCGACGACCTGCCCCGGGCGCCCCCGGGGGGGCTCCGGAA

6361 CACCACCAGCGTGCGCAGCTACCTGCCCAACACGGTGACCGACGCACTGCGGGGGAGCGG
     GTGGTGGTCGCACGCGTCGATGGACGGGTTGTGCCACTGGCTGCGTGACGCCCCCTCGCC
                 ^
                 6372
                 FSP1

6421 GGCGTGGGGGCTGCTGCTGCGCCGCGTGGGCGACGACGTGCTGGTTCACCTGCTGGCACG
     CCGCACCCCCGACGACGACGCGGCGCACCCGCTGCTGCACGACCAAGTGGACGACCGTGC

6481 CTGCGCGCTCTTTGTGCTGGTGGCTCCCAGCTGCGCCTACCAGGTGTGCGGGCCGCCGCT
     GACGCGCGAGAAACACGACCACCGAGGGTCGACGCGGATGGTCCACACGCCCGGCGGCGA

6541 GTACCAGCTCGGCGCTGCCACTCAGGCCCGGCCCCCGCCACACGCTAGTGGACCCCGAAG
     CATGGTCGAGCCGCGACGGTGAGTCCGGGCCGGGGGCGGTGTGCGATCACCTGGGGCTTC

6601 GCGTCTGGGATGCGAACGGGCCTGGAACCATAGCGTCAGGGAGGCCGGGGTCCCCCTGGG
     CGCAGACCCTACGCTTGCCCGGACCTTGGTATCGCAGTCCCTCCGGCCCCAGGGGGACCC

6661 CCTGCCAGCCCCGGGTGCGAGGAGGCGCGGGGGCAGTGCCAGCCGAAGTCTGCCGTTGCC
     GGACGGTCGGGGCCCACGCTCCTCCGCGCCCCCGTCACGGTCGGCTTCAGACGGCAACGG

6721 CAAGAGGCCCAGGCGTGGCGCTGCCCCTGAGCCGGAGCGGACGCCCGTTGGGCAGGGGTC
     GTTCTCCGGGTCCGCACCGCGACGGGGACTCGGCCTCGCCTGCGGGCAACCCGTCCCCAG

6781 CTGGGCCCACCCGGGCAGGACGCGTGGACCGAGTGACCGTGGTTTCTGTGTGGTGTCACC
     GACCCGGGTGGGCCCGTCCTGCGCACCTGGCTCACTGGCACCAAAGACACACCACAGTGG

6841 TGCCAGACCCGCCGAAGAAGCCACCTCTTTGGAGGGTGCGCTCTCTGGCACGCGCCACTC
     ACGGTCTGGGCGGCTTCTTCGGTGGAGAAACCTCCCACGCGAGAGACCGTGCGCGGTGAG

6901 CCACCCATCCGTGGGCCGCCAGCACCACGCGGGCCCCCCATCCACATCGCGGCCACCACG
     GGTGGGTAGGCACCCGGCGGTCGTGGTGCGCCCGGGGGGTAGGTGTAGCGCCGGTGGTGC
```

FIG. 21D

```
6961 TCCCTGGGACACGCCTTGTCCCCCGGTGTACGCCGAGACCAAGCACTTCCTCTACTCCTC
     AGGGACCCTGTGCGGAACAGGGGGCCACATGCGGCTCTGGTTCGTGAAGGAGATGAGGAG

7021 AGGCGACAAGGAGCAGCTGCGGCCCTCCTTCCTACTCAGCTCTCTGAGGCCCAGCCTGAC
     TCCGCTGTTCCTCGTCGACGCCGGGAGGAAGGATGAGTCGAGAGACTCCGGGTCGGACTG

7081 TGGCGCTCGGAGGCTCGTGGAGACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGGAC
     ACCGCGAGCCTCCGAGCACCTCTGGTAGAAAGACCCAAGGTCCGGGACCTACGGTCCCTG

7141 TCCCCGCAGGTTGCCCCGCCTGCCCCAGCGCTACTGGCAAATGCGGCCCCTGTTTCTGGA
     AGGGGCGTCCAACGGGGCGGACGGGGTCGCGATGACCGTTTACGCCGGGGACAAAGACCT
                                ^
                                7167
                                ECO47III

7201 GCTGCTTGGGAACCACGCGCAGTGCCCCTACGGGGTGCTCCTCAAGACGCACTGCCCGCT
     CGACGAACCCTTGGTGCGCGTCACGGGGATGCCCCACGAGGAGTTCTGCGTGACGGGCGA

7261 GCGAGCTGCGGTCACCCCAGCAGCCGGTGTCTGTGCCCGGGAGAAGCCCCAGGGCTCTGT
     CGCTCGACGCCAGTGGGGTCGTCGGCCACAGACACGGGCCCTCTTCGGGGTCCCGAGACA

7321 GGCGGCCCCCGAGGAGGAGGACACAGACCCCCGTCGCCTGGTGCAGCTGCTCCGCCAGCA
     CCGCCGGGGGCTCCTCCTCCTGTGTCTGGGGGCAGCGGACCACGTCGACGAGGCGGTCGT

7381 CAGCAGCCCCTGGCAGGTGTACGGCTTCGTGCGGGCCTGCCTGCGCCGGCTGGTGCCCCC
     GTCGTCGGGGACCGTCCACATGCCGAAGCACGCCCGGACGGACGCGGCCGACCACGGGGG

7441 AGGCCTCTGGGGCTCCAGGCACAACGAACGCCGCTTCCTCAGGAACACCAAGAAGTTCAT
     TCCGGAGACCCCGAGGTCCGTGTTGCTTGCGGCGAAGGAGTCCTTGTGGTTCTTCAAGTA

7501 CTCCCTGGGGAAGCATGCCAAGCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCG
     GAGGGACCCCTTCGTACGGTTCGAGAGCGACGTCCTCGACTGCACCTTCTACTCGCACGC
                                  **********
7561 GGACTGCGCTTGGCTGCGCAGGAGCCCAGGTGAGGAGGTGGTGGCCGTCGAGGGCCCAGG
     CCTGACGCGAACCGACGCGTCCTCGGGTCCACTCCTCCACCACCGGCAGCTCCCGGGTCC
                 ^
                 7575
                 FSP1

                                                    Intron2
     ************************************************************
7621 CCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCCCTGGTCCTCCT
     GGGGTCTCGACTTACGTCATCCCCGAGTCTTTTCCCCCGTCCGTCTCGGGACCAGGAGGA

     ************************************************************
7681 GTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTCGAGTGGACACGGTG
     CAGAGGTAGCAGTGCACCCGTGTGCACCGAAAAGCGAGTCCTGCAGCTCACCTGTGCCAC

     **>
7741 ATCGAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGGTCATA
     TAGCTCCAGCTGAGATCTCCTAGGGGCCCATGGCTCGAGCTTAAGCATTAGTACCAGTAT
          ^
          7747
          SAL1
```

FIG. 21E

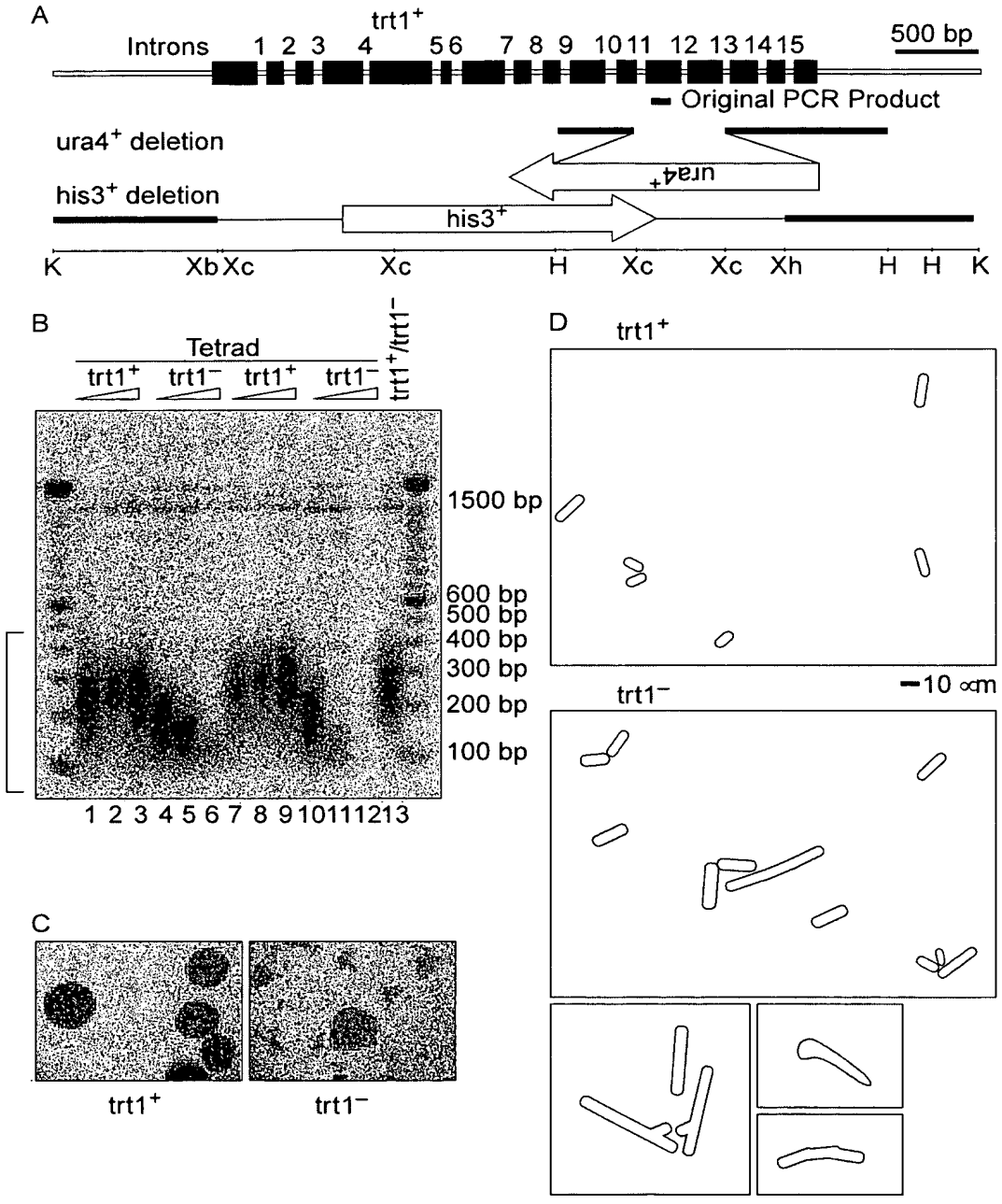
A
trt1+
Introns 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
500 bp
Original PCR Product
ura4+ deletion
his3+ deletion
ura4+
his3+
K Xb Xc Xc H Xc Xc Xh H H K
B
Tetrad
trt1+ trt1− trt1+ trt1−
trt1+/trt1−
1500 bp
600 bp
500 bp
400 bp
300 bp
200 bp
100 bp
1 2 3 4 5 6 7 8 9 10 11 12 13
C
trt1+
trt1−
D
trt1+
trt1−
10 μm

FIG. 22

gccaagttcctgcactggctgatgagtgtgtacgtcgtcgagctgctcaggtctttcttt
tatgtcacggagaccacgtttcaaaagaacaggctctttttctaccggaagagtgtctgg
agcaagttgcaaagcattggaatcagacagcacttgaagagggtgcagctgcgggacgtg
tcggaagcagaggtcaggcagcatcgggaagccaggcccgccctgctgacgtccagactc
cgcttcatccccaagcctgacgggctgcggccgattgtgaacatggactacgtcgtggga
gccagaacgttccgcagagaaaagagggccgagcgtctcacctcgagggtgaaggcactg
ttcagcgtgctcaactacgagcgggcgcg

FIG. 23

TCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGAG
ACCAGCCCGCTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCC
AGCAGTGGCCTCTTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATC
AGGGGCAAGTC

FIG. 24

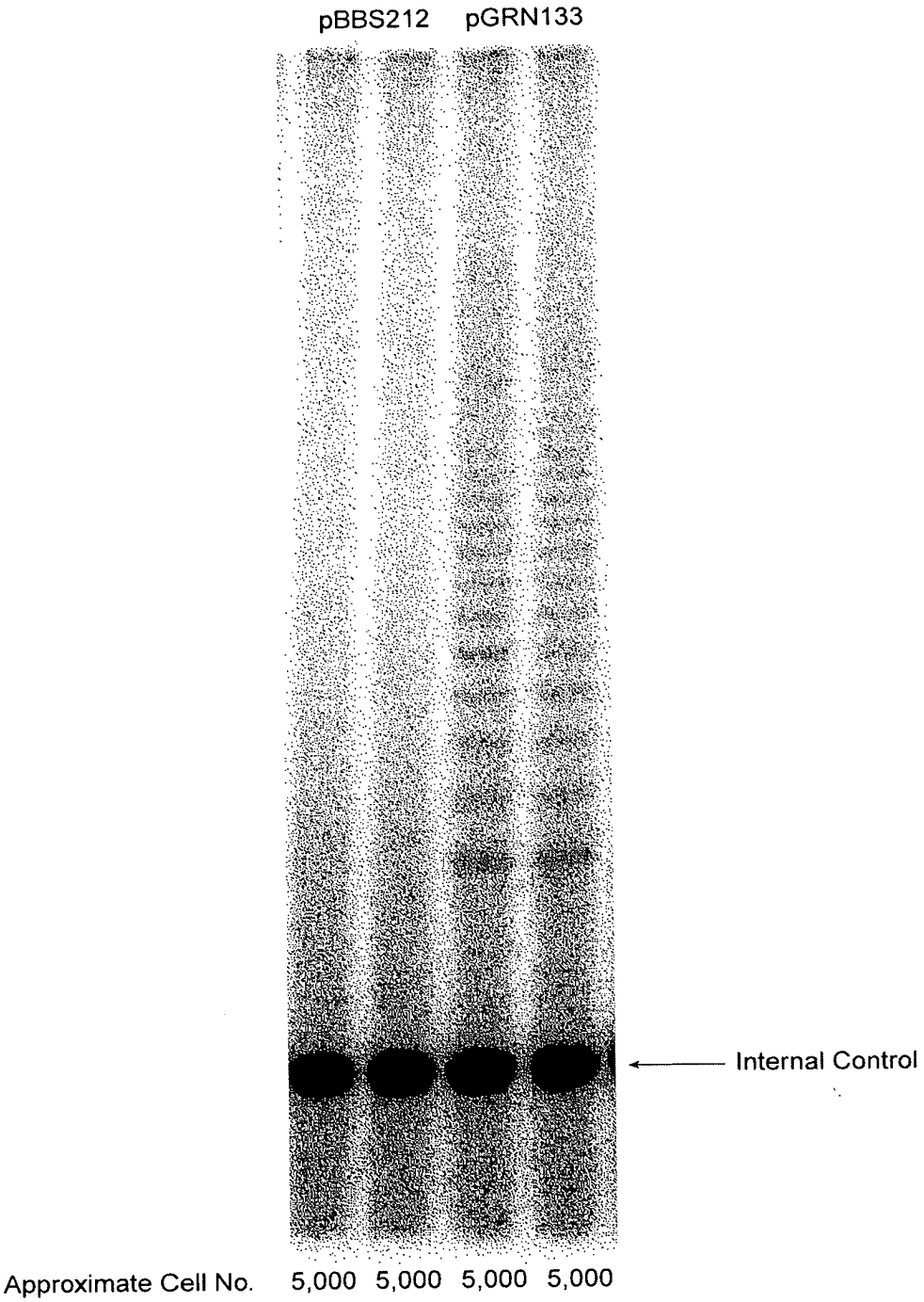
pBBS212
pGRN133
Internal Control
Approximate Cell No.
5,000
5,000
5,000
5,000

FIG. 25